Vol. 44 JANUARY, 1953 No. 1

The Journal of Educational Psychology

UNIVERSITY OF MICHIGAN
MAR 5 1953
PERIODICAL READING ROOM

desk

CONTENTS

Some Effects of Bilingualism upon the Intelligence Test Performance of Puerto Rican Children in New York City 1
ANNE ANASTASI AND FERNANDO A. CORDOVA

An Experimental Comparison of the Stability of Social Acceptability Scores Obtained with the Partial-rank-order and the Paired-comparison Scales 20
SAM L. WITRYOL AND GEORGE G. THOMPSON

Relationships between Rorschach Performance and Student-teaching 31
MARTHA HESSEL PAGE AND ROBERT M. W. TRAVERS

Learning—What Kind of Animal? 41
H. H. REMMERS

A Comparison between Children Who Have Moved from School to School with Those Who Have Been in Continuous Residence on Various Factors of Adjustment. 50
N. M. DOWNIE

A Simplified Item Analysis Card for High-school and College Instructors 54
WILLIAM C. BUDD

Book Reviews 57

Published monthly except June to September

$7.00 a year in the United States, Pan American Countries and the Philippines; Canada, $7.20; other countries, $7.40. Single issues $1.20

WARWICK & YORK, INC.
BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF Educational Psychology

Established 1910

BOARD OF EDITORS*

STEPHEN M. COREY, *Chairman*
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

WM. CLARK TROW
University of Michigan

* Appointed by the American Psychological Association.

EDITORIAL CONSULTANTS

JACK W. DUNLAP
KARL J. HOLZINGER
HAROLD E. JONES
H. H. REMMERS
PERCIVAL M. SYMONDS
MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. The JOURNAL will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

Manuscripts and communications regarding editorial matters should be addressed to the Chairman of the Board of Editors, Stephen M. Corey, Teachers College, Columbia University, New York 27, N. Y.

Books and other materials for review, and business communications should be addressed to THE JOURNAL OF EDUCATIONAL PSYCHOLOGY, Warwick & York, Publishers, 10 E. Centre St., Baltimore 2, Md.

Manuscripts should be typed and double-spaced throughout, including quotations, footnotes, and references. In order to attain

(*Continued on Inside Back Cover*)

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 44 January, 1953 Number 1

SOME EFFECTS OF BILINGUALISM UPON THE INTELLIGENCE TEST PERFORMANCE OF PUERTO RICAN CHILDREN IN NEW YORK CITY[1]

ANNE ANASTASI and FERNANDO A. CORDOVA
Graduate School, Fordham University

Bilingualism is generally recognized as a serious difficulty in the comparative psychological testing of many groups. Although not restricted to this country, the problem is especially cogent in America, with its variety of foreign-speaking subcultures. Many studies on American Indians and on European and Asiatic immigrant groups in America have compared the test performance of bilingual children with that of monolingual, English-speaking children within the same national or racial groups, as well as with the American norms (cf. *1*, pp. 717–725; *3*; *4*; *29*). Such investigations have consistently demonstrated that the substitution of a non-verbal for a verbal test reduces the inferiority of the bilinguals. In some instances, the inferiority disappears completely with the use of the non-verbal test. The relative position of the groups may even be reversed, the bilingual excelling the monolingual on such a test. The latter result is especially likely to occur when the non-verbal test is of the performance type (cf, e.g., *15*).

The interpretation of such findings on bilingualism is complicated by a number of factors. In the first place, it cannot be assumed that verbal and non-verbal tests measure the same functions. The specific abilities called into play by these two types of tests may, and probably do, differ. Do bilinguals score higher on performance than on verbal tests only because their

[1] The present paper is based in part upon an M.A. Dissertation submitted by the junior author to the Department of Psychology, Fordham University (cf. *12*).

insufficient mastery of English handicaps them on the latter tests? Or does the difference in score reflect inequalities in the relative strength of verbal and spatial abilities in the particular group? The bilingual group under investigation may have been reared in a culture which selectively fosters the development of spatial aptitude (or some other ability or combination of abilities involved in the performance test), while discouraging the development of verbal skills.

A second and closely related point to consider is that bilingualism is often correlated with other differences in cultural background which may influence test performance. The bilingual is likely to differ from the American normative population in his general information, work habits, interests, and motivation. The degree of bilingualism, moreover, is probably related to the amount of deviation in these other cultural factors. There is evidence, too, that in certain situations more personal maladjustment occurs among bilinguals than among monolinguals (*30*). To be sure, such maladjustment appears to result primarily from the cultural conflict confronting a second generation immigrant group, but the bilingualism makes the conflict more acute, since it serves as a symbol and constant reminder of such conflict.

A third complicating factor is to be found in the possible influence of bilingualism upon rapport in the testing situation. The bilingual may score higher when tested in one than when tested in the other of his two languages. In a study of Spanish-speaking school children in Arizona, for example, the children obtained higher IQ's on a non-verbal test when the oral instructions were given in Spanish than when they were given in English (*24*). Such a difference may result from a more complete mastery of the one language. But it could also arise, wholly or in part, from a more favorable emotional response toward the examiner who speaks the language with which the individual identifies himself more closely. There is evidence to suggest that both Negro and white children perform better on intelligence tests when the examiner is a member of their own race than when he is a member of a different race (*6*). Like skin color, language may serve as a reduced cue for group identification and it may influence the examiner-subject relationship in a similar manner.

Finally, it should be noted that bilingualism itself may be of

more than one type. Whether or not bilingualism constitutes a handicap, as well as the extent of such a handicap, depends upon the way in which the two languages have been learned (*26*, *31*). In immigrant families, the child often learns one language at home and another at school. The result is a sort of 'linguistic bifurcation,' whereby one language develops in one set of situations and the other in another set. Mastery of both languages is thus limited. It is not the interference between the two languages so much as the restriction in the learning of each to certain areas that leads to handicap. In such cases, the extent of the child's vocabulary as well as other aspects of his linguistic development will be inferior in both languages (*5*, *28*). By contrast, the individual who learns to express himself in all types of situations in at least one language will probably suffer no handicap from learning a second language. Such a situation might be described as 'bilingual parallelism,' since the second language provides a parallel means of expression in some or all situations, depending upon the thoroughness of its mastery.

The present investigation is concerned with the rôle of bilingualism in the intelligence test performance of Puerto Rican children in New York City. The native language of the Puerto Rican is Spanish, although English is taught in the Puerto Rican schools. Unfortunately, an ill-advised educational policy with regard to the way in which English was introduced into the Puerto Rican schools has served only to make many Puerto Ricans 'illiterate in two languages,' and has prejudiced them against English, which they blame for their educational difficulties and confusions (*23*, p. 12). The dark-skinned Puerto Rican migrant, moreover, is encouraged to remain Spanish-speaking, since as a foreign-speaking Negro he tends to enjoy higher status in the United States than does the native American Negro (*23*, p. 87). Such a hybrid Puerto Rican prefers to be identified as a 'Latino' rather than as a Negro. For these and other reasons, the Puerto Rican migrant makes little use of what English he knows. Consequently, Puerto Rican children in New York City are virtually monolingual until school entrance, having had little or no opportunity to learn English in their own home or community. This persisting linguistic barrier is a handicap in the child's school adjustment and in his educational progress. The language problem is also undoubtedly a major reason for the

scarcity of psychological data on this large and rapidly growing immigrant group.

A recently completed sociological survey by Mills, Senior, and Goldsen (*23*) provides some information on the cultural background, socio-economic level, living conditions, interests, and outlook of New York City Puerto Ricans. Strong in-group feeling, low occupational level, and a widespread pessimism regarding the chances of improving their status were characteristic of the sample of 1113 respondents interviewed in this survey. Similarly, an analysis of 188 Puerto Rican first admissions to all New York State mental hospitals during a single year indicated low educational and socio-economic status and a relatively high proportion of cases with subnormal intelligence (*22*).

Only two investigations have been reported in which psychological tests were administered to Puerto Ricans within continental United States, both having been conducted in New York City. The earlier of the two was carried out by Dunklin (*16*) in 1934–1935. Thirty-five Puerto Rican public school children, enrolled in a special section of the first grade taught entirely in Spanish, were given a Spanish translation of the Pintner-Cunningham Primary Test, the Pintner Non-Language Primary Test, the Pintner-Paterson Performance Scale, and another performance test developed by the author. On the two individual performance tests and on the Pintner Non-Language Test, the median IQ of the Puerto Rican children was virtually equal to the American norms, but on the Pintner-Cunningham it was only 73.

The other psychological study of New York City Puerto Ricans was that of Armstrong, Achilles, and Sacks (*2*), reported in 1935. The Army Individual Performance Test was given to 240 Puerto Rican public school children in grades four to six, most of whom were between the ages of nine and fourteen. In addition, the Otis Test of General Ability for grades four to eight was administered in English to 129 of the fifth- and sixth-grade children. As a 'control group,' the authors employed a sampling of over 400 nine- to fourteen-year-old public school children drawn from Manhattan and Westchester County. On both the performance and the verbal test, the mean scores of the Puerto Ricans were lower than those of the control group, the critical ratios of these differences being 7.099 and 9.504, respectively. From these findings, the authors drew rather sweeping conclusions regarding

the inferiority of Puerto Rican intelligence, their report arousing considerable controversy (cf., e.g., *11*). Chief among the difficulties in the way of interpreting the results of this study may be mentioned language handicap, problems of rapport and test motivation, and lack of comparability of Puerto Rican and control groups, especially with reference to socio-economic level.

It may be of interest to consider also some results obtained from the testing of Puerto Rican children outside of continental United States. Porteus (*27*) reports mean Stanford-Binet and Porteus Maze IQ's of 71.72 and 74.87, respectively, for a group of 429 Puerto Rican children tested in Hawaii. The subjects were all clinic cases, brought to the Psychological Clinic of the University of Hawaii, and included delinquent, educationally retarded, mentally defective, and dependent children, from the lower half of the socio-economic scale. Within this group, the Puerto Rican children scored lower than any of the other national samples tested. There is no indication of how the language problem was handled, if at all. Moreover, the possibility of differential selection of various national subgroups within the clinic population needs to be taken into account. An extensive and better controlled investigation was conducted in Puerto Rico under the direction of The International Institute of Teachers College, Columbia University, and reported in 1926 (*25*). In addition to island-wide testing with school achievement tests given both in English and in Spanish, this study included the administration of the Pintner Non-Language Mental Ability Test to 1000 children in grades three to eight. In the latter test, the Puerto Rican children excelled the American norms in grades three, four, and five, but fell below these norms in the next three grades.

PROCEDURE

In the present investigation, Puerto Rican children in the upper three grades of a New York City elementary school were tested with the Cattell 'Culture Free' Test, Forms 2A and 2B, designed for ages eight to twelve and unselected adults (*7, 8, 9, 10*). This is a non-verbal test, all items being perceptual or spatial. Cattell maintains, moreover, that the oral instructions for this test can be translated without altering the validity of the test (*9*, p. 3). It should be added that for the present purpose no assumption

has been made that this test is actually 'culture free.' No test can be completely 'culture free,' nor even 'culture constant,' since the content of any test will tend to favor one or another culture. The elimination of specific culturally limited information from a test is only a partial and superficial solution. Each culture stimulates the development of certain abilities and interests, and inhibits others. The resulting psychological differences will inevitably be reflected in test performance, as in any other behavior of individuals reared in diverse cultural settings. In the Cattell test, for example, the items consist almost exclusively of abstract geometric forms and patterns; and the test is, of course, of the paper-and-pencil variety. Individuals having relatively little interest and practice in intellectual games and those who have developed antagonistic or discouraged attitudes toward purely academic activities would quite probably be at a disadvantage on such a test, which provides little practical appeal.

The plan of the present experiment was to administer Form A in English to one half of the subjects (Group 1), and Form A in Spanish to the other half (Group 2). After an interval of two weeks, Form B was given with a reversal of the two languages, Group 1 now receiving the instructions in Spanish, and Group 2 in English. The basic experimental design was thus a 2×2 latin square, as shown below:

	Test Session	
	I	II
Group 1	Form A: English	Form B: Spanish
Group 2	Form A: Spanish	Form B: English

Both forms of the test were employed with each language in order to balance out any possible differences in the difficulty level of Forms A and B. Although Cattell reports (*9*, p. 3) that the two forms are equal in difficulty, such equivalence might not hold for the present population. The study was therefore set up in such a way as to make the assumption of equality of forms unnecessary.

The Spanish translations of the instructions were prepared by the junior author, who is himself a bilingual Puerto Rican, and were independently checked by a Spanish-speaking reader. All tests were likewise administered by the junior author, with the assistance of classroom teachers. The children were tested in

groups containing from thirty-five to forty-nine subjects. Preliminary introductory remarks, as well as replies to subjects' queries, were made in the same language as the instructions during any one test session.

SUBJECTS

The subjects tested in the present study included 176 boys and girls in the sixth, seventh, and eighth grades of a parochial school. The school was situated in the principal area of residence of Puerto Ricans in New York City, where the two earlier studies on New York City Puerto Ricans, described in the first part of this paper, had also been conducted. Of the total group of 176, thirty-two were omitted owing to absence from one of the testing sessions, seven because they were not Puerto Rican, and three because of zero score on one or both forms of the test. The three last-mentioned subjects had made no effort to put any marks on the test items. The remaining 134 cases were employed in determining the reliability coefficients of each form of the test. A further reduction in number was made prior to carrying out the analysis of variance, in order to have an equal number of cases in each subgroup. For this purpose, the number of cases retained was 108, consisting of twenty-seven boys and twenty-seven girls in Group 1 and the same numbers in Group 2. Within each school grade, individual subjects were assigned to Groups 1 and 2 by the experimenter in such a way as to equate the two groups as nearly as was practicable in relevant background characteristics.

At the end of the second testing session, each subject filled out a personal data form. This questionnaire was written in English, but the examiner read each question aloud in both English and Spanish, explaining how it should be answered and giving individual attention to subjects who raised questions or required assistance. The results of this questionnaire served as the basis for describing the sample under investigation, as well as for determining how closely Groups 1 and 2 had been equated. The questionnaire data are reported for only the 108 subjects used in the analysis of variance, since it is on this group that the major conclusions of the study are based.

In age, the group ranged from eleven to fifteen years, with a mean of thirteen. Nearly all parents were born in Puerto Rico,

only one or two fathers or mothers in each subgroup having been born in other Spanish-speaking countries or in the United States. Among the children, nearly half were born in Puerto Rico, the rest in New York City; nearly all had lived in New York City over three years. Since Puerto Ricans make frequent return trips to their native island, questions were also inserted to cover the number and duration of such visits. Nineteen of the 108 children reported such trips, but only four of these had remained in Puerto Rico over one year during their visits. As in other surveys of Puerto Rican migrants, the occupational level of the present group was very low. When classified according to the Goodenough and Anderson (*17*) occupational scale, none of the parents fell into the professional, semi-professional, clerical and skilled trades, or farmer categories. The majority were engaged in semi-skilled and slightly skilled occupations.

Nine of the eighteen items in the questionnaire were concerned with the extent of the subject's bilingualism. These items were selected and adapted from a bilingualism scale developed by Hoffman (*20*). In replying to each question, the subjects checked one of the following alternatives: (a) always in Spanish; (b) more in Spanish than English; (c) about half and half; (d) more in English than Spanish; (e) always in English. Most children reported that they spoke Spanish and English about equally often with their families, a larger number clustering at the all-Spanish than at the all-English end of the scale. The rest of the family, however, most often spoke Spanish among themselves. Reading by the family scatters widely over the scale, as does letter-writing, although Spanish predominates in the latter. The children themselves employ predominantly English in their reading and writing, a fact which undoubtedly reflects the influence of the school. English language movies are the more frequently attended, although Spanish movies are well represented in the group. Radio listening shows a more even distribution of Spanish and English, with a slight predominance of English. Finally, a clear majority of the children indicated that they 'think' in both languages; and as between the two ends of the scale, a larger number fell at the English than at the Spanish end of the scale on this item.

The four subgroups proved to be closely equated in all of the above background variables, with only the following minor excep-

tions: the Group 2 boys included a slight excess of New York-born cases, were somewhat superior in parental occupational level, and reported a little more frequent use of English. Because of longer residence in this country, the families of these boys may thus have made a more effective adjustment to the American environment, both economically and culturally. That this subgroup difference was, however, too slight to have appreciably influenced the results of the present study is suggested by the fact that in over-all test performance the Group 1 boys excelled those in Group 2. The only clear sex difference noted in the questionnaire data is a greater tendency for the girls to report attendance at Spanish movies. This difference may result in part from the greater restriction of the girls' activities in Puerto Rican families, which would delay their assimilation to the American culture. But it may also be related to the nature of Spanish films, which are more often of the 'romantic' type and would thus appeal more to girls than to boys at the pre-adolescent and adolescent level.

RESULTS

Test reliability.—Split-half reliability coefficients were computed separately for Forms A and B, in both the English and Spanish versions. These coefficients were obtained by finding each subject's score on the odd items (a), the even items (b), and the total test (t), and applying the following formula (*13*, *19*):

$$r_{11} = 2\left(1 - \frac{\sigma_a^2 + \sigma_b^2}{\sigma_t^2}\right).$$

The reliability coefficients, together with the number of subjects on whom they were computed, will be found in Table 1. As a

TABLE 1.—SPLIT-HALF RELIABILITY COEFFICIENTS

Test Form and Version	N	r_{11}
A—English	61	.92
A—Spanish	73	.84
B—English	73	.88
B—Spanish	61	.88

check on the applicability of an odd-even reliability technique to the present test, the degree to which the scores depended upon speed was determined. This information was also of direct

interest in so far as there are cultural differences in the emphasis placed upon speed. Such differences would be reflected in performance on a speed test.

As an index of speed, the following measure developed by Cronbach and Warrington (*14*) was computed for each subtest within each form and version of the test:

$$\frac{\sigma^2_{t(a)}\,\Sigma X^2_{\lambda p}}{N\sigma_x^2}$$

In this formula, σ_x^2 is the total variance of the test scores, and $\sigma^2_{t(a)}$ is the variance of the number of items attempted. To find $\Sigma X^2_{\lambda p}$, each person's score on the last two items attempted is averaged, the average is squared, and this quantity is then summed for all persons. Adjustments are also made for any persons who may have completed the test before time is called. Essentially, this index is based upon the extent of individual differences in number of items completed, as well as upon performance on the last two items attempted. Thus an individual who failed the last two items attempted would probably not raise his score appreciably if given time to try more, and harder, items. An estimate of the lower bound of the reliability coefficient of a speed test can be found by subtracting this index from the obtained reliability coefficient.

TABLE 2.—INDEX* SHOWING THE CONTRIBUTION OF SPEED TO SCORES ON EACH SUBTEST

Subtest	Form A English	Form A Spanish	Form B English	Form B Spanish
1	.0513	.6068	.0358	.0278
2	.0093	.3114	.0011	.0021
3	.0308	.0681	.0000	.0218
4	.0038	.0000	.0000	.0003

* Computed by formula of Cronbach and Warrington (14, p. 175).

The results of the speed analysis are given in Table 2. On the whole, the values reported in this table are quite low, all but two falling between .0000 and .0681. The two exceptions are subtests 1 and 2 of Form A administered in Spanish. Scores on these two subtests were evidently influenced to a considerable extent by speed, probably because most subjects were slow in starting after receiving the instructions. For the present pur-

pose, only total scores on all subtests are of interest. It is likely that the reliability coefficient of .84 obtained for Form A—Spanish has been spuriously raised to some extent through the contribution of speed to scores on two of its four subtests. In all other forms, however, it is apparent that speed played a negligible rôle.

Analysis of variance.—The mean scores of boys and girls in Groups 1 and 2 during each of the two testing sessions are reported in Table 3. It will be recalled that in Group 1, the

TABLE 3.—SUBGROUP MEANS OF TEST SCORES

Testing Session	Group 1 (A—English; B—Spanish)		Group 2 (A—Spanish; B—English)	
	Boys	Girls	Boys	Girls
Session 1: Form A	20.93	15.93	15.70	18.30
Session 2: Form B	24.52	22.19	21.89	25.07

first testing session was conducted in English and the second in Spanish, while the reverse order was followed in Group 2. In order to evaluate the contribution of language, session, order, and sex to test performance, the data were submitted to an analysis of variance. The type of analysis employed is essentially the same as that described by Grant (*18*) for a 2×2 latin square, except for the addition of the sex variable and the computation of the interaction of sex with each of the other three variables. The interactions among language, session, and order are confounded and lost, as in all 2×2 latin squares.

In order to determine the applicability of analysis of variance to the present data, the eight subgroups were tested for homogeneity of variance by the L_1-test (cf. *21*, pp. 82–86). The obtained value of L_1 is .9687; with $f = 26$ and $k = 8$, we find that $P > .05$. Hence the variances may be treated as homogeneous.

The results of the analysis of variance are summarized in Table 4. In computing the F-ratios, the subject $\times$ session interaction was used as the error term; in one case, a second F-ratio

TABLE 4.—ANALYSIS OF VARIANCE

Source of Variation	df	Sums of Squares	Variance Estimate	F-ratio
1) Order	1	22.686	22.686	1.152
2) Session	1	1756.741	1756.741	89.202*
3) Language	1	32.667	32.667	1.659
4) Sex	1	8.167	8.167	.415
5) Session × Sex	1	35.852	35.852	1.820
6) Language × Sex	1	14.518	14.518	.737
7) Order × Sex	1	580.166	580.166	29.459*
				9.908*
8) Subjects (within order—within sex)	104	6090.074	58.558	2.973*
9) Error (subjects × session)	104	2048.222	19.694	
10) Total	215	10589.093		

* Significant at .01 level.

was found with another error term, as will be explained below. It will be noted that of the four principal factors investigated, viz., order, session, language, and sex, only session yielded a significant F-ratio. This ratio, 89.202, is considerably greater than the value of 6.90 required for significance at the .01 level, with $1/104$ df. Such a finding indicates a large and significant practice effect, all groups scoring much higher in the second session regardless of language used. Of the 108 subjects, eighty-five improved on the retest. Since this group was taking its first psychological test, the rôle of test sophistication and rapport is strongly suggested by these results. Whereas in the test manual Cattell allows one point for test sophistication when Form B follows Form A (*9*, p. 8), the present group showed a mean rise of 5.71 points. Cattell elsewhere (*8*, p. 157) cites evidence to suggest that gains between Forms A and B tend to be greater for individuals with higher IQ's. In the present group, however, over-all level of performance, or IQ, was low in terms of the Cattell norms, while retest gains were relatively high. This is what would be expected when initial scores are spuriously lowered by failure to understand directions, poor motivation, and the like.

The language in which the test was administered had no significant effect upon performance. Translating the directions evidently offers no solution to the testing of these subjects. The bilingualism encountered in this group appears to be of the bifurcated variety, the children's mastery of either language being restricted and inadequate. The teachers report that these children show a reluctance to speak English. Yet in Spanish, the majority are illiterate. In connection with the present testing, a number of the children told their teachers that the Spanish used by the examiner was 'too correct' for them. Nor could this objection refer to regional differences in Spanish idiom or accent, since the examiner was a native Puerto Rican. Another interesting illustration of the linguistic bifurcation of these subjects is the difficulty which many experienced with the word 'subrayar' (to underline), which they had more often encountered in the English-speaking school situation than in the Spanish-speaking home setting. It should also be noted that the linguistic bifurcation provides a further reason for the improvement upon a retest, since by the second session all subjects had heard similar instructions in both languages, and thus had a better opportunity to complement their inadequate understanding of either language. It is clear that in future testing of Puerto Rican migrants, the best procedure is either to use exclusively non-language tests or to present the instructions in both languages.

There was no significant sex difference in over-all performance, but the sex × order interaction was significant at a high level of confidence. Such a significant interaction could result either from the differential effect of order upon the two sexes, or from the fact that the subjects differed widely among themselves, since each order was employed with a different sampling of subjects. As a check on these alternative explanations, a second F-ratio was computed for the sex × order interaction, using the subject variance as the error term.[1] This F-ratio also proved

[1] It will be noted that the subject variance is itself significant, when tested against subject × session interaction as the error term. This simply means that, despite marked improvement from first to second session, subjects tended to maintain their relative status in the group. In other words, differences among individuals were significant and persisted throughout the experiment.

to be significant at the .01 level. Hence the significant interaction could not be attributed to sampling fluctuations which might have made the Group 1 girls inferior or the Group 1 boys superior by chance. The obtained interaction effect indicates that the testing order which began with English favored the boys, while that beginning with Spanish favored the girls. Such a sex difference probably reflects differences in the degree of acculturation of the two sexes. Several questionnaire items suggested that the boys were more highly Americanized than the girls. This difference may be due in part to the slight excess of New York born boys in the total sample. Another and more general explanation is to be found in the greater freedom traditionally allowed to boys in Puerto Rican families, while girls are more closely restricted to the family circle.

It is interesting to note that the language × sex interaction was not significant. Neither sex did significantly better in one language than in the other. The Spanish instructions proved as difficult as the English for all groups, probably because of the unfamiliarity of key words resulting from the subjects' linguistic bifurcation. At the same time, the initial use of Spanish or English by the examiner may have influenced rapport, an influence which would carry over to the second test session. The boys may have felt more coöperative and at ease with an initially English-speaking examiner, the girls with an initially Spanish-speaking examiner. The sex difference would thus be a matter of attitude rather than linguistic comprehension.

Comparison with normative sample.—The over-all performance of the present group fell considerably below the norms reported in the test manual (*9*, p. 9). For this comparison, the Form B scores of the total sample of 108 children were converted to standard score IQ's. The Form B norms, as given by Cattell, are based on the assumption that Form A was taken prior to Form B, as was done by the present subjects. Moreover, Cattell recommends that, with subjects lacking in test sophistication, Form B scores are a more satisfactory measure of performance, Form A serving as a practice test (*8*, p. 156). The median standard score IQ obtained by the present sample was 70. The range extends from two top IQ's of 124 down to three scores indicating chance performance, i.e., raw scores below 9.2 (*8*, p.

157). The distribution of IQ's is significantly skewed, with a piling up of scores at the lower end.

In interpreting these standard score IQ's, it should be noted that they are based on a standard deviation of 24 points, as contrasted with the more familiar 16- or 15-point σ in use in such scales as the Stanford-Binet and the WISC. Thus the median IQ of 70 is 1.25σ below the norm of 100, and would correspond to an IQ of approximately 80 or 81 in terms of the more familiar units. The nature of the normative sample must also be taken into account. The norms were obtained on 3297 British and American children, ranging in age from six to seventeen. Cattell reports that "a substantial fraction of the (normative) population . . . was taken either from two Midwestern university towns of population about twenty-five thousand each, or from a British industrial city of population about two-hundred-forty thousand" (*8*, p. 156). Both Midwestern towns were rated above average in cultural and social status in the survey conducted by E. L. Thorndike (cf. *8*, *32*). Parental occupation is not reported by Cattell, but it is almost certain that the normative sample is of much higher socio-economic level than the present group. The low socio-economic status of the Puerto Rican children, as indicated by parental occupation, is undoubtedly one factor to consider in evaluating their poor test performance. Their language handicap, limiting their mastery of both Spanish and English, is another important factor.

Even more conspicuous as a reason for the inferior test perfomance of the present group is the children's attitude toward the testing. A large number of factors, including lack of test sophistication, little or no motivation to excel in a competitive intellectual situation, and lack of interest in the relatively abstract and academic content of the test contributed to this attitude. The characteristic reaction to the testing was a mild confusion, followed by amusement and indifference. Such attitudes, moreover, are closely related to the poor emotional adjustment which this group makes to the school. The children were described by their teachers as 'unambitious,' many just sitting in the classroom without understanding what goes on. Their initial school experiences, involving sudden placement in an exclusively English-speaking environment at a time when they knew almost

no English, seem to have produced in these children a sort of 'psychological insulation' to what goes on in school. The passive and unresponsive habits thus established have remained their characteristic reaction to school. A solution of the language problem early in the child's school career would thus seem to be an essential first step for the proper education of these children. Not only test performance, but also the general intellectual development which the tests are designed to gauge, are seriously handicapped by the attitudes and intellectual habits resulting from the child's early linguistic confusion.

It may be of interest to reëxamine the findings of Dunklin (*16*) and of Armstrong et al. (*2*) in the light of the present results. The subjects in the Dunklin study were first-grade children and were attending a special section taught in Spanish. Thus these children had not yet developed the passive, unresponsive attitudes characteristic of the older school children. Rapport was also better because the teacher, as well as the examiner, spoke to them in the more familiar Spanish. Under these conditions, Dunklin found no inferiority to the American norms on three of the tests. The inferiority on the Spanish translation of the Pintner-Cunningham may be attributed to the fact that the Spanish which the children had acquired in their own homes failed to provide the type of vocabulary required to understand psychological test instructions. In the Armstrong et al. study, which was conducted on fourth- to sixth-grade children, consistently poor scores were obtained on both verbal and performance tests. Although no language was employed on the performance test, the inferiority of the Puerto Rican children on this test was probably due partly to poor rapport in the testing situation and partly to the cumulative effects of four to six years of unsuitable schooling.

SUMMARY

The Cattell Culture Free Intelligence Test, Forms 2A and 2B, was administered to 176 Puerto Rican children in grades six to eight of a parochial school in the Spanish Harlem area of New York City. One half of the group received the test instructions in English during the first testing session (Form A) and in Spanish during the second session (Form B); the order of the languages was reversed for the other half of the group. The

split-half reliability of Forms A and B in the English and Spanish versions ranged from .84 to .92. Speed played a negligible part in the scores obtained.

An analysis of variance was conducted on 108 of the subjects, including twenty-seven boys and twenty-seven girls in each of the two language-order subgroups. Significant F-ratios were found for two variables, subjects and session, and for the interaction of order × sex. The most conspicuous finding was the marked improvement from first to second testing session, regardless of language. Although there was no over-all sex difference in score, the girls performed better when the testing order was Spanish-English, the boys when it was English-Spanish. This order × sex interaction was attributed principally to rapport, the more highly Americanized boys responding more favorably to an initially English-speaking examiner, while the more restricted and less acculturized girls achieved better rapport with an initially Spanish-speaking examiner.

The over-all performance of the present group fell considerably below the test norms reported by Cattell. Among the reasons for such a discrepancy are the very low socio-economic level of the Puerto Rican children, their bilingualism which makes them deficient in both languages, their extreme lack of test sophistication, and their poor emotional adjustment to the school situation. In so far as this maladjustment itself appears to have arisen from the children's severe language handicap during their initial school experiences, a solution of the language problem would seem to be a necessary first step for the effective education of migrant Puerto Rican children.

REFERENCES

1) A. Anastasi and J. P. Foley, Jr. *Differential Psychology* (Second Ed.). New York: Macmillan, 1949, pp. 894.

2) C. P. Armstrong, E. M. Achilles and M. J. Sacks. "Reactions of Puerto Rican children in New York City to psychological tests." *Rep. Spec. Comm. on Immigration and Naturalization, New York State Chamber of Commerce*, 1935, pp. 9.

3) S. Arsenian. *Bilingualism and Mental Development.* Teachers College, Columbia Univ., Contrib. to Educ., 1937, No. 712, pp. 164.

4) ———. "Bilingualism in the post-war world." *Psychol. Bull.*, 1945, 42, 65–86.

5) E. M. Barke and D. E. P. Williams. "A further study of the

comparative intelligence of children in certain bilingual and monoglot schools in South Wales." *Brit. J. educ. Psychol.*, 1938, 8, 63–77.

6) H. G. Canady. "The effect of 'rapport' on the IQ: a new approach to the problem of racial psychology." *J. Negro Educ.*, 1936, 5, 209–219.

7) R. B. Cattell. "A culture-free intelligence test: I." *J. educ. Psychol.*, 1940, 31, 161–179.

8) ———. "Classical and standard score IQ standardization of the I.P.A.T. Culture-Free Intelligence Scale 2." *J. consult. Psychol.*, 1951, 15, 154–159.

9) R. B. Cattell and A. K. S. Cattell. *Culture Free Intelligence Test. Scale 2. Handbook.* Champaign, Ill.: I.P.A.T., 1949, pp. 9.

10) R. B. Cattell, S. N. Feingold and S. B. Sarason. "A culture-free intelligence test: II. Evaluation of cultural influences on test performance." *J. educ. Psychol.*, 1941, 32, 61–100.

11) P. A. Cebollero. *Reactions of Puerto Rican Children in New York City to Psychological Tests. An Analysis of the Study by Armstrong, Achilles, and Sacks of the Same Name.* San Juan, P. R.: The Puerto Rico School Review, 1936, pp. 11.

12) F. A. Cordova. *A Comparison of the Performance of a Bilingual Group on a "Culture Free" Test Administered in English and in Spanish.* Unpub. M.A. Disser., Fordham Univ., 1951.

13) L. J. Cronbach. "Coefficient Alpha and the internal structure of tests." *Psychometr.*, 1951, 16, 297–334.

14) L. J. Cronbach and W. G. Warrington. "Time-limit tests: estimating their reliability and degree of speeding." *Psychometr.*, 1951, 16, 167–188.

15) N. T. Darcy. "The effect of bilingualism upon the measurement of the intelligence of children of preschool age." *J. educ. Psychol.*, 1946, 37, 21–44.

16) L. D. Dunklin. *A Study of the Intelligence of Some Recent Puerto Rican Immigrant Children in a First Grade in a New York City School.* Unpub. M.A. Disser., Teachers College, Columbia Univ., 1935.

17) F. L. Goodenough and J. E. Anderson. *Experimental Child Study.* New York: Century, 1931, pp. 546.

18) D. A. Grant. "The statistical analysis of a frequent experimental design." *Amer. J. Psychol.*, 1949, 62, 119–122.

19) L. Guttman. "A basis for analyzing test-retest reliability." *Psychometr.*, 1945, 10, 255–282.

20) M. N. H. Hoffman. *The Measurement of Bilingual Background.* Teachers College, Columbia Univ., Contrib. Educ., 1934, No. 623, pp. 75.

21) P. O. Johnson. *Statistical Methods in Research.* New York: Prentice-Hall, 1949, pp. 377.

22) B. Malzberg. "Mental disease among Puerto Ricans in New York State." *Psychiat. Quart. Suppl.*, 1948, 22, 207–218.

23) C. W. Mills, C. Senior and R. K. Goldsen. *The Puerto Rican Journey.* New York: Harper, 1950, pp. 238.

24) A. J. Mitchell. "The effect of bilingualism in the measurement of intelligence." *Elem. Sch. J.*, 1937, 38, 29–37.

25) P. Monroe. *A Survey of the Public Educational System of Porto Rico.* New York: Bur. Publ., Teachers College, Columbia Univ., 1926, pp. 453.

26) R. Pintner and S. Arsenian. "The relation of bilingualism to verbal intelligence and to school adjustment." *J. educ. Res.*, 1937, 31, 255–263.

27) S. D. Porteus. "Racial group differences in mentality." *Tabulae Biologicae*, 1939, 18, 66–75.

28) M. E. Smith. "Measurement of vocabularies of young bilingual children in both of the languages used." *J. genet. Psychol.*, 1949, 74, 305–310.

29) D. T. Spoerl. *The Adjustment at College of Students Who Were Bilingual in Childhood.* Unpub. Ph.D. Disser., Clark Univ., 1942.

30) ———. "Bilinguality and emotional adjustment." *J. abn. soc. Psychol.*, 1943, 38, 37–57.

31) W. A. Stark. "The effect of bilingualism on general intelligence: an investigation carried out in certain Dublin primary schools." *Brit. J. educ. Psychol.*, 1940, 10, 78–79.

32) E. L. Thorndike. *Your City.* New York: Harcourt, Brace, 1939, pp. 204.

AN EXPERIMENTAL COMPARISON OF THE STABILITY OF SOCIAL ACCEPTABILITY SCORES OBTAINED WITH THE PARTIAL-RANK-ORDER AND THE PAIRED-COMPARISON SCALES

SAM L. WITRYOL

Department of Psychology, The University of Connecticut

and

GEORGE G. THOMPSON

School of Education, Syracuse University

This investigation was designed to compare the relative stability of social acceptability scores obtained with the partial-rank-order and the paired-comparison scales when administered to school children. The partial-rank-order sociometric approach has been popularly associated with the creative efforts of Moreno (*2*) and his associates. It has been very widely employed by a host of investigators for research purposes and by an increasingly large number of teachers, psychologists, sociologists, and others interested in its diagnostic possibilities. In marked contrast, an intensive review of the literature (*10*) has yielded only a few examples of the employment of the paired-comparison sociometric technique.

The term 'stability' is employed in this study as a combined measure of test-retest reliability and social consistency. It ordinarily implies the test-retest method with emphasis upon errors of measurement. Contemporary usage is clarified by Cronbach's (*1*, p. 69) definition: "The coefficient of stability shows the extent to which scores on the particular test items are stable over a period of time. It indicates whether a sample of behavior taken at one time is typical of behavior at other times." Stability coefficients for a particular test are, of course, influenced by a number of variables: size and composition of the experimental population, the nature of the testing instrument, dynamic social change, and so forth. The effects of such variables on the stability of sociometric scores were discussed in some detail by the present writers in a recent monograph (*10*) in which several studies of the stability of partial-rank-order sociometric scales were analyzed. Few investigations of the stability of the paired-

comparison sociometric scale have been conducted. The present investigation permits a direct comparison of the relative stability of the two instruments, since they were administered to the same populations on alternate days under similar conditions.

Practical and theoretical features of the partial-rank-order approach to social acceptability have been reported in companion studies by Thompson, Bligh, Powell, and Witryol (*7*, *8*). The primary concern in the present investigation was to make an objective comparison of the relative stability of sociometric scores obtained with the two scales when they were administered on three different occasions to four sixth-grade classes. A secondary purpose was an attempt to determine the comparability of scores obtained with the two scales. Do they measure different or highly similar aspects of social status? The feasibility of employing the paired-comparison sociometric instrument in the classroom was also examined.

EXPERIMENTAL PROCEDURE

The sociometric instruments.—The partial-rank-order sociometric scale administered to the school children was adapted from an instrument employed by Northway (*3*) and her students in the Toronto studies. The subjects were instructed to write in order of preference three choices of associates in response to four questions:

> 1) Suppose you were to move to another classroom. Which boys or girls from this classroom would you like best to go with you?
>
> 2) Which boys or girls of the classroom would you like to play with during recess?
>
> 3) What do you like to do best in school?_____ Which boys or girls of the classroom would you like to do this with you?
>
> 4) What do you like to do best out of school?_____ Which boys or girls of the classroom would you like to do this with you?

A child's social status score was the sum of all the choices he received on all four criteria. Ranks were not weighted. The questions were assumed to imply a broad enough range of behavior so that the sum total of the responses might reflect a general index

of social acceptability rather than associational preferences specific to one type of classroom activity.

The paired-comparison sociometric questionnaire was constructed in the following manner. The name of each subject in a particular classroom was paired with the name of each of his classmates. The total possible pairings of the pupils in each group are defined by the formula, $\frac{n(n-1)}{2}$, where n equals the number of pupils. The pairings were mechanically arranged to minimize position effects. This means that there was a maximum separation between two pairings containing the same name in each, and that each name occurred approximately an equal number of times on the left as on the right side of a given pairing. Ross (*6*) and Wherry (*9*) have developed rational and empirical procedures for dealing with this problem of ordering the pairs.

The pupils were asked to underline on a mimeographed form " . . . the name of the person that you like the better . . . " in each pairing. The index of social acceptability was the total number of times an individual was chosen by his classmates in all the pairings where his name occurred.

The experimental population and design.—The subjects were pupils in four different sixth-grade classrooms in three different schools. The number of pupils in these groups ranged from nineteen to twenty-five. It may be noted in the subsequent tables of results that stability and interrelational measures are based on varying N's. These adjustments were necessary because of absences. Status scores of children who were absent when the partial-rank-order scale was administered were eliminated from stability calculations, on the assumption that children appear to operate in part on an 'out of sight, out of mind' basis when recording choices on this instrument. These eliminations probably tended to increase the stability of the partial-rank-order scores. No corrections for absences were made in the data obtained with the paired-comparison instrument because this approach to the determination of social status in the classroom is not so seriously affected by absences.

Both sociometric questionnaires were administered on three different occasions to the pupils in the four sixth-grade groups. After initial presentation, the two instruments were readministered one week later and, again, five weeks from the time of the

first testing period. On each of the three occasions the two tests were given one day apart in counter-balanced order. In two of the classrooms, the partial-rank-order form was presented first on all three occasions; in the other two classrooms, the paired-comparison form was administered first on all three occasions.

RESULTS

Two types of correlational analysis were made of the social acceptability scores obtained with the two scales. Stability coefficients were calculated from the social status scores derived in each classroom for each of the two scales over intervals of one, four, and five weeks. Thus, six Pearson product-moment stability coefficients were calculated for each classroom from sociometric scores obtained by the partial-rank-order and the paired-comparison methods. This yielded a total of twenty-four coefficients for the four classroom groups. Intercorrelations were also obtained between the scores on both scales administered approximately at the same time on each of three different occasions. The total of these product-moment intercorrelations for the four classroom groups was twelve.

The stability coefficients presented in Table 1 will be examined first. These results may be summarized as follows:

1) All the paired-comparison coefficients are above .90. The range is .904 to .987 with a median of approximately .96. The stability coefficients obtained when the paired-comparison sociometric scale was applied to sixth-grade children are substantially high and consistently greater than those obtained with the partial-rank-order scale.

2) Over the same temporal periods, the partial-rank-order sociometric stability coefficients computed from the same classroom groups as in (1), above, range from .595 to .964 with a median of approximately .75. However, four of the twelve coefficients are .903 or higher. The stability of the partial-rank-order scores fluctuated considerably from classroom to classroom. For example, all the coefficients computed from Classroom 'D' are above .90, but none of the correlations calculated from scores in Classroom 'C' is above .80.

3) There appears to be a slight tendency for the magnitudes of the stability coefficients derived from both scales to decrease

with an increase of the time intervals over one-, four-, and five-week periods. In Classroom 'A' the apparent decrease in stability with time is quite pronounced for the partial-rank-order scale.

TABLE 1.—STABILITY COEFFICIENTS OF SOCIOMETRIC SCORES OBTAINED ON THREE DIFFERENT OCCASIONS IN FOUR SIXTH-GRADE CLASSROOMS

Classroom	Sociometric Instrument*	Time Interval Between Administrations					
		One Week		Four Weeks		Five Weeks	
		r	N	r	N	r	N
'A'	Paired-Comparison	.975	25	.927	25	.938	25
	Partial-Rank-Order	.941	22	.658	22	.595	21
'B'	Paired-Comparison	.987	20	.980	20	.963	20
	Partial-Rank-Order	.711	14	.828	15	.638	18
'C'	Paired-Comparison	.962	19	.921	19	.904	19
	Partial-Rank-Order	.693	18	.690	18	.805	17
'D'	Paired-Comparison	.971	21	.969	21	.947	21
	Partial-Rank-Order	.964	18	.942	17	.903	17

* On each of the three occasions the two tests were separated by one day; the partial-rank-order scale was administered the first day in two of the classrooms, and the paired-comparison scale was administered the first day in the other two classrooms.

The results of the intercorrelations between the two sociometric scales appear in Table 2. The range of the twelve intercorrelations between the two scales is .363 to .895 with a median of approximately .70. The intercorrelations between the two scales are uniformly high in Classroom 'D' (.889, .895, .857) where the stability coefficients of both sociometric scales are relatively high (all above .90). Conversely, in those instances where the stability coefficients are considerably lower—and this is characteristic of the results obtained with the partial-rank-order scale

TABLE 2.—INTERCORRELATIONS BETWEEN PARTIAL-RANK-ORDER AND PAIRED-COMPARISON SOCIOMETRIC SCORES OBTAINED ON THREE DIFFERENT OCCASIONS

Classroom	Different Administrations of Both Tests					
	First		Second*		Third†	
	r	N	r	N	r	N
'A'	.822	23	.795	23	.643	23
'B'	.417	15	.770	19	.621	19
'C'	.673	19	.363	18	.541	18
'D'	.889	18	.895	18	.857	20

* One week after first administration of both tests.
† Five weeks after first administration of both tests.

in this study—the intercorrelations between the two scales are substantially lower.

DISCUSSION AND CONCLUSIONS

It may be concluded that the paired-comparison sociometric approach is a somewhat more stable measure than the partial-rank-order scale in the assessment of social status within sixth-grade classrooms over time intervals of one to five weeks. This is consistent with the findings of non-comparative investigations reviewed in a recent monograph by the present writers (*10*). This conclusion would also be drawn on a purely theoretical basis, i.e., the paired-comparison procedure is a longer 'test' in the sense that more responses are elicited from each child.

The stability coefficients obtained in the present study with the paired-comparison scale are sufficiently high to satisfy the conventional criteria for reliability of measurement. This finding is consonant with results obtained by previous investigators. In their experimental examinations of various approaches to sociometric measurement, Eng and French took the position (*5*, p. 368): " . . . Because of the rigor of the method of paired comparisons it has been suggested that it serve as a criterion of

validity for other psychological scaling methods . . . ." The possibility remains, however, that the somewhat lower stability coefficients of the partial-rank-order scale can be explained in part by a greater sensitivity to the varying psychological-social factors in a dynamic social setting. In other words, errors of measurement are confounded with 'real' social change in any stability approximations of this type. Nevertheless, it seems apparent that the 'forced-choice' characteristic of the paired-comparison method provides certain advantages with respect to reliability of measurement. This technique forces each member of a group to make psychological-social decisions about every other member of the group in every possible combination of pairs. This makes the scale more sensitive to the social status of those individuals who occupy intermediate positions on the underlying and theoretically 'real' continuum. Thompson and Powell (*8*) have discussed this aspect of sociometric procedure in another type of 'forced-choice' situation, the rating scale. The partial-rank-order scale tends to focus the attention of the subjects upon a few of their best friends, or upon the socially rejected children, if that end of the continuum is called for in a particular questionnaire. This approach tends to neglect those individuals falling in the middle part of the social-status distribution. Northway has concluded from test-retest data obtained with the partial-rank-order method (*4*, p. 59): " . . . The least accepted individuals retain . . . status, the most accepted retain theirs and the fluctuation occurs in middle groups."

The influence of errors of measurement on data secured with the partial-rank-order scale is discussed in the Thompson-Powell study (*8*). The populations in that experiment were similar to those employed in the present investigation, i.e., sixth-grade school children. However, the social groupings were larger, more than thirty in each classroom. Consequently (on the basis of larger N's), the stability coefficients obtained in that study over the same temporal intervals with the same partial-rank-order instrument were somewhat higher. The range of the obtained stability coefficients was .84 to .92 with a median of approximately .86. The range in the present experiment, employing smaller N's in each group, was .60 to .96 with a median of approximately .75. Since the partial-rank-order sociometric scores appear to be somewhat more stable in the larger groups,

one is tempted to infer that the errors of measurement are reduced simply as a function of the larger N's. However, there are certain reasons for suspecting that this may not be an entirely satisfactory answer.

For example, the magnitudes of the obtained correlations between the partial-rank-order and paired-comparison social status scores in this study might suggest that both scales are measuring a great deal in common. This might suggest that their communality is attenuated only by errors of measurement. However, it is obvious that the intercorrelations of the two scales corrected for known errors of measurement would not approximate unity in some of the classrooms. It may be that each scale is measuring somewhat different facets of social-status structure. This conclusion is not unreasonable when one examines the types of decisions a subject is required to make in responding to the two experimental situations.

The subject makes a limited number of choices in several social situations on the type of partial-rank-order scale employed for this study. These choices may reflect momentary popularity, recent association, stable friendship, or a combination of these factors. The paired-comparison scale forces a comparative rating on some general social-attraction basis of everyone in the group, regardless of familiarity or intensity of feeling. The individual's social status in a group is approximated in more general terms by his score on the paired-comparison scale. This general social-status position is highly stable over time intervals as long as five weeks. Stability coefficients of partial-rank-order measures of social status vary considerably from classroom to classroom, but the median of the coefficients is reasonably satisfactory for most purposes. In a given classroom the limited number of social preferences expressed on this scale may make these choices more subject to change in psychological meaning. They may, in general, reflect more personal preferences than those in the paired-comparison approach where total scores may mask the more intimate and enduring social relationships. Varying stability of the partial-rank-order sociometric scores from classroom to classroom (as compared to the consistently high stability of the paired-comparison scores) may be attributable, then, to:

1) Larger errors of measurement due to the relative 'shorter' length of the test.

2) The intimate nature of the choices, making them more subject to recent situational factors.

3) The consequences of some changes of preference within a small number of possible choices—especially in a small social grouping.

4) The more specific definitions of the situations within which choices are to be made—permitting the influence of a larger variety of social forces which may function differentially on different occasions.

If one considers both scales as measuring somewhat the same 'common core' of social-status, a decision as to which scale to employ will be determined largely by stability requirements in a particular project and by practical convenience. The paired-comparison approach would be chosen on the basis of its superior stability of measurement. However, this technique is extremely time-consuming to construct and to score. Furthermore, it does not yield a manageable sociogram of over-all social interrelationships in a group.

The following conclusions about the relative merits of the two scales have been made as first approximations:

1) The paired-comparison sociometric scale is highly reliable and yields more stable scores than the partial-rank-order instrument.

2) Although the partial-rank-order scale yields scores that are reasonably stable, the stability coefficients may vary considerably from one social grouping to another.

3) Correlations between social-status scores obtained with the two scales demonstrate a high degree of communality between the social-psychological variables being measured.

4) It is suggested that the paired-comparison scale yields a relatively more general index of social status and that the partial-rank-order technique reflects a relatively more personal or intimate social preference for classroom associates. The latter may be more sensitive to the effects of 'real' changes in social status of a situational type.

5) The selection of one scale or the other should probably not be based entirely on either the criteria of reliability or convenience; a further and perhaps more crucial consideration in choosing between the two scales hinges on the definition of social acceptability most relevant to a particular experimental or

diagnostic objective. Objective data on which such a decision might be made are unfortunately not available.

SUMMARY

The purpose of this study was to compare experimentally the stability of social acceptability scores obtained with the partial-rank-order and paired-comparison scales. Both scales were administered to the same school children in four different sixth-grade classrooms on three different occasions. Both scales were readministered on a second occasion, four weeks later, and on a third occasion, five weeks from the time of initial presentation. The classrooms ranged in size from nineteen to twenty-five pupils.

Twelve stability coefficients calculated from the sociometric scores obtained by the method of paired-comparisons over time intervals of one to five weeks ranged from .904 to .987 with a median of approximately .96. Under comparable experimental conditions with the same subjects, the partial-rank-order scale yielded twelve stability coefficients, ranging from .595 to .964 with a median of approximately .75. The twelve correlation coefficients obtained between the scores on the two sociometric scales ranged from .363 to .895 with a median of approximately .70.

It is concluded that the paired-comparison scale is a highly reliable instrument and that social acceptability scores derived with this technique are more stable than those obtained with the partial-rank-order approach. The partial-rank-order scale of social status yields reasonably stable scores, but the scores vary considerably from classroom to classroom. This variability of the stability of scores in different social groups suggests the possibility that the partial-rank-order scale may reflect a somewhat different type of social status than the paired-comparison instrument. On the other hand, a fairly high degree of communality between the two scales in social-psychological factors measured is suggested by the substantial intercorrelation between scores secured with the two instruments. It is suggested that the scores obtained with the paired-comparison scale are relatively more general measures of social status, whereas partial-rank-order scores probably reflect more personal and situational factors. These latter social preferences may be more sensitive to dynamic social changes in a given group structure than are the

more general measures of social status. This hypothesis appears sufficiently plausible to merit further investigation.

REFERENCES

1) L. J. Cronbach. *Essentials of Psychological Testing.* New York: Harper, 1949.

2) J. L. Moreno. *Who Shall Survive?* Washington: Nervous and Mental Disease Publishing Co., 1934.

3) M. L. Northway. "Instructions for using the sociometric test." *Sociometry,* 1946, 9, 242–248.

4) M. L. Northway, E. B. Frankel, and R. Potashin. *Personality and Sociometric Status.* Sociometric Monogr., 1947, No. 11.

5) E. Eng and R. L. French. "The determination of sociometric status." *Sociometry,* 1948, 11, 368–371.

6) R. T. Ross. "Optimum orders for the presentation of pairs in the method of paired comparisons." *J. educ. Psychol.,* 1934, 25, 375–382.

7) G. G. Thompson, H. F. Bligh, and S. L. Witryol. "A critical examination of several methods of determining levels of social status." *J. soc. Psychol.,* 1951, 33, 13–32.

8) G. G. Thompson and M. Powell. "An investigation of the rating-scale approach to the measurement of social status." *Educ. psychol. Measmt.,* 1951, 11, 440–455.

9) R. J. Wherry. "Orders for the presentation of pairs in the method of paired comparisons." *J. exper. Psychol.,* 1938, 23, 651–660.

10) S. L. Witryol and G. G. Thompson. *A Critical Review of the Stability of Social Acceptability Scores Obtained with the Partial-rank-order and Paired-comparison Scales.* Genet. psychol. Monogr. (In press.)

RELATIONSHIPS BETWEEN RORSCHACH PERFORMANCE AND STUDENT-TEACHING[1]

MARTHA HESSEL PAGE

Onieda Child Guidance Center, Utica, N. Y.

and

ROBERT M. W. TRAVERS

Personnel Research Laboratory
Lackland Air Force Base
San Antonio, Texas

Numerous studies have been conducted in which there has been an attempt to measure some aspect of the teacher's adjustment and some aspect of teacher effectiveness. Such studies have attempted to correlate the two measures in order to determine the relationship between them. Because of the complicated nature and the different definitions assigned to both 'adjustment' and 'effectiveness,' the statistical measures obtained have not as yet succeeded in clarifying the relationships between the two aspects of teacher behavior. Therefore, this study was concerned with a simpler problem. Was there a relationship between the patterns of behavior derived from supervisors' descriptions of student-teachers and these students' Rorschachs? The aim was to determine if there were personality patterns which had certain consistent relationships with the criterion used—supervisors' descriptions of the characteristics of students of education in a student-teaching situation.

PROBLEM

The problem selected for preliminary exploration was that of studying the behavior of the student-teacher in a complex testing situation and then searching for relationships between this behavior and behavior described by the supervisor. The Rorschach Test was selected for this preliminary study for several reasons. First, in the administration of the Rorschach it is possible to study the behavior of the individual in making adjustments to a complex problem-solving situation, and thus it

[1] This paper is a report of part of a research project sponsored and developed by the Division of Teacher Education of the City College of New York.

was anticipated that there might be some relationship between some of these behaviors and those manifested in the complex problem-solving situation faced in student-teaching. Second, the Rorschach Test is by far the best developed instrument for the study of personality by means other than verbal report or interview.

It did not seem reasonable that the behavior of most teachers could be adequately described in terms of a few simple dimensions for which ratings might be made on a graphic rating scale. Therefore, after several meetings with the supervisors in which the problem of description was discussed, a form was devised in which the supervisor was merely asked to describe the outstanding characteristics of the student as a teacher in a classroom. These descriptions were regarded not as criteria of effectiveness but as reflections of the supervisors' value judgments and concepts of desirable and undesirable student-teacher behavior.

PROCEDURE

Individual Rorschachs were administered to students of education at a municipal college of a large city. These were individuals who planned to begin student-teaching during the following semester. Students were asked by the Rorschach worker and the faculty to coöperate in a study of the kinds of persons entering teaching. The confidential nature of the study was stressed and students were urged to make appointments for testing. The department, however, was active in making arrangements for testing and encouraging students to coöperate.

The individual Rorschach Test was administered with verbatim recording. No tests were scored until all Rorschach testing was completed in order to reduce any bias in scoring. Scoring was based primarily on Munroe's modification of Klopfer's system so that the revised Munroe Check List might be used (*5*). Interviews which dealt with the students' attitudes and aspirations were held. A student interest questionnaire, which obtained information on interests and home background, was also administered.

POPULATION

Sixty-four individuals were tested and interviewed by one worker. There were thirty-nine students in elementary educa-

tion, four of them men and thirty-five women. Of the twenty-five in secondary education, nine were men and sixteen women. Mean age for those in elementary education was 22.9 years with a SD of 4.5 and a range of 20 to 42. Mean age for those in secondary education was 21.9 years with a SD of 2.23 and a range of 19 to 27. The thirty-one students in elementary education for whom A.C.E. scores were available had a mean score of 113.06, with a SD of 18.4 and a range of 70 to 146. This mean differed at the .05 level from that of the twenty-second secondary education students for whom A.C.E. scores were available. The mean score of the latter group was 124.41 with a SD of 17.8 and a range of 93 to 159. However, cumulative grade point averages for the first four semesters' work in college revealed only a trend toward greater achievement for the secondary education students.

A majority of the subjects were first or second generation Americans from the less formally educated and less prosperous segments of the population and were among the first in their families to attend college. They were socially mobile, using teaching not only as an outlet for their energies and as a means of earning a living and obtaining security, but also as a step upward in the social scale. They had been liberally exposed to a test-minded, competitive system of education throughout their lives.

The faculty of the education department believes in a mental hygiene, pupil-centered philosophy of education. Every effort is made by the faculty to study the individual student for the purpose of eliminating those whose inadequacies would seriously handicap them as teachers and for the additional purpose of helping undesirable but remediable characteristics. The students appeared to be on guard in the Rorschach situation and seemed to be anxious to appear 'normal' on the test.

RESULTS

The individual Rorschach records of the education students were characterized for the most part by a great deal of restraint and a lack of high-level organizational activity. The median number of responses for the thirty-nine elementary education students was 26.5, with a range of 14 to 125, while for the twenty-five secondary education students the median was 30 and the range 14 to 131. According to the literature high level groups

like these tend to give a greater median number of responses although 30 is presumably most frequent. Administration of the test, because of the brief Rorschach protocols obtained, usually took less than an hour. The students expressed nervousness and made many comments during the course of the test, such as "is this right?," "with a stretch of the imagination it could be," "are you supposed to see more than one thing?" "do other people see this?" Thirty-five students of the sixty-four had to be asked on Card I if they saw more than one thing. Three individuals rejected cards.

Although it had originally been intended to write predictions from the Rorschachs to be matched up with the supervisors' descriptions, this project was abandoned because of the brevity and stereotypy of the majority of protocols obtained. Recent investigators (*3*, *4*) have urged caution in the application of standard Rorschach interpretations to protocols of non-pathological individuals. The individual's mental set caused by situational factors like those mentioned above on the Rorschach performance may warp the results in the direction of unrepresentativeness. Furthermore, to write a prediction of behavior in student-teaching which was more than a series of glossy generalizations would be a difficult task in view of the heterogeneous nature of student-teaching positions.

Because of the different demands on elementary and secondary school teachers, the two groups were treated separately. Supervisors' descriptions were available for thirty-one elementary education students. These descriptions were easily classified into two categories: those which described behavior which was desirable and those which indicated undesirable behavior. Desirable behavior patterns included initiative, poise, self-confidence, creativeness, responsibility, good planning, dependability, good interpersonal relationships with pupils and authority figures, insight into children's behavior, eagerness for criticism and discussion of problems, flexibility, capacity for growth and coöperativeness. An example of such behavior is given below:

"S is a happy, relaxed prospective teacher. She is warm and friendly with children and has deep insight into their problems. She is especially effective with children on the individual or small group basis and acquired a great deal of skill in handling the group situation. She uses resources intelligently. She plans

well and includes children in the planning whenever it is possible to do so. She gets along well with supervisors, coöperating teachers and the staff in general. She faces her own problems honestly and with a nice sense of humor and does something about them."

An undesirable behavior pattern is given below: (Other undesirable patterns included difficulty with authority, lack of responsibility, autocratic methods with children, etc.)

"S is a quiet, friendly girl who is inclined to be shy and self-conscious. Does not have too much confidence in herself. Is inclined to pout over her inadequacies. Her emotions are plainly shown by her expressions. She is kind, considerate and likes children and they respond to her, although in one situation, they did not readily accept her and she was very upset. She also displayed some jealousy in the situation in regards to another student. This student had been there for some time and was secure with the children and the teacher. S felt she was on the 'outside.' Her relationship on the supervisory level showed this need for personal acceptance. S will be an average teacher and will operate effectively when she feels that she has the support and acceptance of her group and her supervisor."

Undesirable patterns did not imply failure, since those in student-teaching were a highly selected group. They seemed to imply that such students were not exhibiting in one area or another the traits and patterns of behavior valued most highly by the supervisors. These values were apparently consistent from one supervisor to the other in the coöperating group in light of the terminology used and the specific incidents cited to illustrate behavior. Supervisors were not responding to adjustment in general, but adjustment to a specific situation—student-teaching. They were giving patterns which they felt to be desirable in terms of their own values.

The two classifications, 'desirable' and 'undesirable' behavior patterns as reported by supervisors, were related to Rorschach findings in various ways. One approach was to use the adjustment score derived from the Revised Munroe Inspection Technique and Check List. This consists of a list of Rorschach elements and patterns, deviations from which are said to indicate emotional disturbances of certain kinds. The number of these deviations can be totalled and a numerical score of adjustment

obtained. Munroe reports that the median number of checks in the group Rorschach for most college and other non-pathological populations is 10. The median number of checks for the elementary education group was 11.3 with a range of 6 to 20. Greater opportunity to score for deviations is afforded in the individual administration. Furthermore, although the students showed much color and shading shock, this was felt to be not so much an indication of basic personality disturbance as of rigidity in the face of the unexpected in a situation where they felt it important not to do anything 'abnormal.' Their responses were increasingly vague, diffuse, and stereotyped on the critical cards, but only in a few instances bizarre.

It seems quite inevitable that genuine attempts to screen out those who are personally unfitted for teaching as happened in this school will produce considerable anxiety in the student body. This anxiety would possibly generalize to a test like the Rorschach, in spite of assurances of secrecy. Therefore, in view of their social mobility, their awareness of the selection procedures in their college and particularly in their department, and their previous competitive experiences with tests, as well as the fact that most of them are in a measure still dependent on their families and working through such dependencies, one would anticipate a considerable amount of tension in the test. This would be attributable partly to their interpretations of the testing situation as a test of their fitness for teaching and partly to their personal conflicts and achievement needs. These students coöperated, because coöperation is encouraged in their department. The extent to which they gave a representative sample of their behavior when they did not feel so much on trial and so at a loss for the 'proper' response posed a problem which should be considered by those using the Rorschach for investigation purposes with a 'coöperating' group.

Cooper and Lewis (*1*) and Cronbach (*2*) failed to find relationships between signs of adjustment on the Munroe Check List and ratings of performance in a classroom. However, to confirm their findings and to demonstrate further that 'adjustment' was too ambiguous a term to use in discussing teacher personality, Check List scores were divided into two categories. 'Normal' adjustment was represented by a score of 10 and below; 'questionable adjustment,' by a score of 11 and above. No significant

relationship was found between scores on the Check List and desirable behavior in this group (Chi-Square = .074). Apparently desirable behavior in student-teaching was not incompatible with a high level of tension on the Rorschach. The clinical concept of adjustment, as represented by the Munroe Check List, has thus been demonstrated by three experiments using different procedures and populations to be of little predictive value.

TABLE I.—RELATIONSHIP BETWEEN RORSCHACH TRIAD AND SUPERVISORS' CHARACTERIZATIONS OF BEHAVIOR PATTERNS FOR 31 ELEMENTARY EDUCATION STUDENT-TEACHERS

	Supervisors' Characterizations		
	Desirable Behavior Pattern	Undesirable Behavior Pattern	Total
Showing Rorschach Triad	10	2	12
Not Showing Rorschach Triad	4	15	19
Total	14	17	31

An approach which was not concerned with 'adjustment' but methods of coping with problems was taken. Desirable behavior seemed to be associated with an emotionally outgoing, rather ambitious and labile orientation to problems, a primary interest being in people and in the environment. Since the conventional concept of adjustment was not a useful one in this group, a search was made for patterns of behavior in the Rorschach which were related to patterns of behavior considered desirable. It soon was found that possession of a triad of patterns in the Rorschach, a ratio of two or more whole responses (W) to one human movement response (M), the weighted sum of color responses (ΣC) greater than human movement responses (M), and color-form plus color responses (CF + C) greater than form-color responses (FC) was associated in this group with desirable behavior as defined by the supervisors. Table I shows the relationships of

possession and non-possession of this pattern to desirable and deviations from desirable behavior.

The value of Chi-Square was computed and found to be 9.14, which is significant at the .01 level of confidence (one degree of freedom). Obviously this does not provide a validation of the triad, since it was derived from the group under study. Furthermore, the factors used to make up this triad have not been found to possess high reliability, especially in brief protocols such as those obtained here. With a different attitude the subject might have a different distribution of responses. In a number of cases the ratios were 'borderline' and additional responses could have altered them.

However, the strong drive, the desire for achievement, the emotional outgoingness, the outward orientation and interest in environment and people, and the lability and suggestibility which are interpretations of the different ratios in the pattern seemed to be characteristics most closely associated with the kinds of behavior described as desirable. An individual with a drive to achieve and to please and with a basic interest in people and the environment would be more likely to profit from the course of training at this college and to show the kinds of behavior which ranked high in the supervisors' opinions. This type of person might be a bit too adaptable and suggestible and might in a different environment adopt behavior calculated to please those in the present and not in the past. It would almost seem as if the introversive individual, even if he were not greatly disturbed emotionally, according to the check list, just did not fit into the desirable pattern as well as did the driving, labile extratensive person. However, further refinements of this technique involving classification of responses and scoring, and cross-validation of the experiment are needed.

At the secondary education level, there seemed to be no patterns of behavior on the Rorschach Test which were consistently associated with any particular pattern of behavior reported by supervisors in student-teaching. Supervisors mentioned the intellectual approach and initial awkwardness with children. However, pupil-centered teaching at this level demands considerable skill in teaching and much experience with children. Furthermore, secondary education student-teachers did not have as much opportunity for prolonged contact with

children as did those in elementary education. Certainly this indicates that personality characteristics which are important in elementary education may be subordinate to situational factors such as the intellectual level of the work, the maturity and greater independence of the students and the shorter time spent with the pupils in secondary education.

SUMMARY AND CONCLUSIONS

1) The individual Rorschach was administered to thirty-nine elementary education and twenty-five secondary education applicants for student-teaching at a municipal college. The aim of the study was to find a relationship between patterns of behavior, as reported by supervisors in a student-teaching situation. The descriptions of behavior were not used as criteria of effectiveness but as reflections of supervisors' value judgments.

2) Supervisors' descriptions of thirty-one elementary education student teachers' outstanding characteristics were divided into two categories: those which consisted of patterns of traits and behaviors regarded as desirable by supervisors and those which consisted of patterns regarded as undesirable. Desirable behavior patterns seemed to be those which might be called extraverted, emotionally outgoing, ambitious and labile.

3) There was no relationship between adjustment scores derived from the Munroe Check List and category of behavior for elementary education students. A high amount of tension was not incompatible with desirable behavior in student-teaching. The inadequacy of the clinical concept of adjustment applied to teaching was pointed out. The factors contributing to the tension of the students in the Rorschach situation such as certain situational factors and the emotional conflicts intrinsic in this late adolescent age group were discussed. The need for caution in using Rorschach results from 'coöperating' groups was emphasized.

4) Certain suggestive relationships were found between a triad of Rorschach ratios and desirable performance. The triad consisted of a W to M ratio of 2 or more W to 1 M, M less than ΣC, and FC less than CF + C. It would seem that the faculty felt that the extraverted, ambitious person exhibited the kind of behavior which was most satisfactory for child-centered, informal teaching. However, the need for further investigations of relia-

bility and for cross-validation in a comparable group of students from that college was pointed out.

5) Among the student-teachers in secondary education no pattern on the Rorschach was found related to any category of performance in student-teaching. The conclusion was that personality characteristics were subordinate to the situational factors at this level of education, such as the age of the pupils, amount of time spent with them, and the intellectual level of the work.

REFERENCES

1) James G. Cooper and Roland B. Lewis. "Quantitative Rorschach Factors in the Evaluation of Teacher Effectiveness," *Journal of Educational Research*, 44: 703–707, May, 1951.

2) Lee J. Cronbach. "Memorandum on the Rorschach Test in Teacher Selection," Washington: State College of Washington, pp. 6, June, 1946.

3) Max. L. Hutt, R. Gibby, E. O. Milton, and K. Pottharst. "The Effect of Varied Experimental 'Sets' upon Rorschach Test Performance," *Journal of Projective Techniques*, 14: 181–187, June, 1950.

4) Abraham S. Luchins. "Situational and Attitudinal Influences on Rorschach Responses." *American Journal of Psychiatry*, 103: 780–784, May, 1947.

5) Ruth L. Munroe. "The Inspection Technique for the Rorschach Protocol," pp. 91–145 in L. Abt and L. Bellak (Eds.) *Projective Psychology*, New York: Alfred A. Knopf, 1950.

LEARNING—WHAT KIND OF ANIMAL?*

H. H. REMMERS

Purdue University

My distinguished predecessor in his presidential address last year presented it under the title: "Learning—They Ain't No Such Animal!"[1] That stimulating and challenging paper, along with the need to give a presidential address myself, provided the stimulus for some further consideration of his thesis from the viewpoint of educational psychology.

From the arcana of the theorizers on learning there have recently issued several questionings, misgivings, and, in the case of English, downright repudiation of the need for or indeed the possibility of a self-consistent, meaningful and useful theory of learning. Let me hasten to say that it is not my purpose, even if I had the time and the ability, to bring order out of chaos in the area of learning theory. Rather I wish for this occasion (1) to present some of the doubts, misgivings and strictures that have been made concerning the matter, (2) to look briefly at the problem semantically and (3) to offer a few examples of the kinds of variables that have been in my judgment neglected as to their importance for learning as the educator generally employs that term.

One of the doubters, Skinner, recently published a paper entitled: "Are Theories of Learning Necessary?"[2] To make his semantic referent of 'theory' clear I quote his definition. He means by it "any explanation of an observed fact which appeals to events taking place somewhere else, at some other level of observation, described in different terms, and measured, if at all, in different dimensions." He believes that "it might be argued that the principal function of learning theory to date has been, not to suggest appropriate research, but to create a false sense of security, an unwarranted satisfaction with the status

* Presidential Address before the Division on Educational Psychology of the American Psychological Association, Washington, D. C., September 1, 1952.

[1] H. B. English. "Learning—they ain't no such animal!" *Jour. edu. Psychol.*, 43: 6, October, 1952, pp. 321–330.

[2] B. F. Skinner. *Psychological Review*, Vol. 57, No. 4, July, 1950, pp. 193–216.

quo." To illustrate: "getting out of a box faster and faster is not learning; it is merely performance. The learning goes on somewhere else, in a different dimensional system. And although the time required depends upon arbitrary conditions, often varies discontinuously, and is subject to reversals of magnitude, we feel sure that the learning process itself is continuous, orderly, and beyond the accidents of measurement. Nothing could better illustrate the use of theory as a refuge from the data" (p. 195).

At this point the semanticist might want to reëxamine critically the concept of theory implied. It differs, it seems to me, in important respects from the statement made by my professor of chemistry: "A hypothesis is a guess, a theory is a good guess, and a law is a guess you can prove." I leave you to wrestle further with the problem.

Skinner's conclusion in any event is one with which I have no quarrel: " . . . it is possible that the most rapid progress toward an understanding of learning may be made by research that is not designed to test theories. An adequate impetus is supplied by the inclination to obtain data showing orderly changes characteristic of the learning process. An acceptable scientific program is to collect data of this sort and to relate them to manipulable variables, selected for study through a common sense exploration of the field" (p. 215).

Too, there will be general acceptance of his further statement: "This does not exclude the possibility of theory in another sense. Beyond the collection of uniform relationships lies the need for a formal representation of the data reduced to a minimal number of terms. A theoretical construction may yield greater generality than any assemblage of facts. But such a construction will not refer to another dimensional system and will not, therefore, fall within our present definition. It will not stand in the way of our search for functional relations because it will arise only after relevant variables have been found and studied. . . . We do not seem to be ready for theory in this sense" (p. 46).

Agreement with this point and elaboration of it is given by Koch[3] with respect to motivational psychology, who notes that psychology seems to be in a state of total disorientation not only

---

[3] Sigmund Koch. "Current status of motivational psychology," *Psychological Review*, Vol. 58, No. 3, May 1951, pp. 147–154.

with reference to the whole body psychological as was the case in the two decades before the World War II, but also in its component parts, as was not true or at least apparent then. This disorientation is nowhere more evident, he believes, than in the field of motivation. Witness such diversity of theory as connoted by names as Freud, McDougall, Lewin, Guthrie, Hull, Morgan, Köhler, E. L. Thorndike, etc.

Learning is in a pre-theory stage of development using 'theory' now to mean self-consistent explanation based on empirical findings. Whether the concept of learning, as English asserts, is a white elephant supported chiefly by hordes of white rats, or whether it will turn out to be a useful and productive cow to provide nourishment and to be properly fitted into the metaphorical phylogenetic psychological series beginning with a pure sensation as the analogue of the one-celled animal, is as yet too early to say. Shifting the figure of speech from biology to physics, we are nearer Thales than Galileo.

When we turn from the "empty superstructure of current theory"[4] to the available empirical data we are impressed, as are Koch and English in the papers already mentioned, with their meagerness and inconclusiveness, as well as the triviality of much of it, at least from the point of view of the teacher who wants to bottom educational practice on sound scientific principles. For a concentrated documentation of this assertion I commend you to Hovland's competent and scholarly chapter on "Human Learning and Retention."[5] The teacher who comes to this summary with a healthy intellectual appetite will find little sustenance for his day-to-day labors to say nothing of longer units of time. It turns out to be a largely meatless bone of an as yet largely unknown animal.

In another paper Koch[6] turns depth-psychologist using the clinician's criterion of a set of private correlations to diagnose the syndrome of attitudinal complexes of psychologists that he believes are widely held and are retarding the progress of theo-

---

[4] Koch, op. cit., p. 150.

[5] Carl I. Hovland. "Human Learning and Retention," Chapter 17 in Stevens' *Handbook of Experimental Psychology*, New York: John Wiley & Sons, Inc., 1951, pp. 613–688.

[6] Sigmund Koch. "Theoretical psychology, 1950: An overview." *Psychological Review*, Vol. 58, No. 4, July, 1951, pp. 295–301.

retical psychology incalculably. Since my own prejudices largely coincide with his I quote his diagnosis.[7]

"(1) The belief that psychology had already accumulated a sufficient number of cold, hard and basic 'empirical facts' to justify attempts at comprehensive theory. The stage was set for an era of Galilean construction, or Newtonian systematization, or, at the very least, Baconian descriptive organization.

"(2) The belief that for areas where cold, hard facts were conspicuously missing, the situation could be rectified by founding theories on areas in which the cold, hard facts were presumably available, and then deriving consequences which would fill in the gaps.

"(3) The belief that typical experimental situations for the study of a limited set of representative behaviors could be identified by *a priori* analysis as the exclusive induction basis for a group of theoretical laws adequate to all behavior.

"(4) The belief that theoretical programs were theories, that hypothetical guesses were laws, that anticipated consequences of hazily limned in points of view were theorems.

"(5) The belief that quantification of theoretical relationships, or their coördination to extant mathematical system, was either immediately possible, or would automatically become so, without further methodological analysis.

"(6) The belief that experimental data collected by members of one's own theoretical tribe could be trusted, while out-group data could not be trusted.

"(7) The belief that rival theoretical positions could be given the decisive *coup de grâce* by critical experiments.

"(8) The belief that, at bottom, any theoretical assumption, no matter how foolhardy the inductive leap (assuming that there is something to leap from), is better than none at all.

"(9) The belief that all will be well, provided one listens to his theoretical father-surrogate, or, more succinctly, the departmental theory of truth.

"If there is anything to be rejected within the pre-war culture that I am trying to describe, it is this still tenacious, attitudinal—or, if you please, delusional—system."

My good friend English, frustrated by three decades of failure to encompass learning in one or a few simple formulations, solves

[7] Koch, op. cit., p. 297.

the problem by " . . . they ain't no such animal—" for psychology as 'pure' science. He would, however, assign the term an honorable place in 'applied' psychology. To quote (without his permission) from a letter from him under date of August 5:

"Of course there is such an animal! It is an applied science animal, a mule if you will, half donkey, half horse. Very useful! Bears the burdens of practical life but sterile for breeding scientific theories."

Koch[8] detects in the present crisis technologists responding in one of two ways. They either believe that the success of their "rule of thumb" procedures can "achieve full fruition without theoretical support" or that "their applied discoveries are, in reality, the foundation material for the psychological theory of the future."

In this dichotomy of 'pure' versus 'applied' (impure) psychology, since my scientific efforts have been highly 'impure,' the paranoid component of my personality structure detects a latent snobbery and/or a semantic confusion. Science as method differs not one whit for the activities subsumed under the two terms, hence on that ground there is no reality in the dichotomy. The only difference then would appear to be in the purposes of the two alleged kinds of scientists. If the activity is for 'useful' ends it is by definition technology, leaving whatever is useless for the 'pure' psychologist. It reminds me of the man who, unable to get along with his wife, decided to separate from her and to divide with her their common property, so he divided the house by giving her the 'outside' of the house and kept the 'inside' for himself.

This tongue-in-cheek observation must not be interpreted to mean an anti-theoretical position. In the long run demonstrably sound practice must inevitably rest upon equally demonstrably sound theory. But such theory, whether provided by the pure psychologist or the technologist, will be of the inductive variety created not full-blown like Athene from the brow of some psychological Zeus in an imagined world, but by slow accretion and coalescence from testing hypotheses of limited scope in a real world, where the scientists of whatever sort speak the same scientific language concerning operationally defined variables.

---

[8] Koch, op. cit., pp. 297–298.

Some twenty-odd years ago John Dewey wrote a little book[9] in which he pointed out that education like medicine must incorporate the results of all of the relevant sciences and that at the time of his writing the only area in which this lesson had been learned was in that branch of mathematics called statistics. Since then social psychology, psychiatry, cultural anthropology and sociology have been developing leads and techniques that need very much to be incorporated in the armamentarium of the educational psychologist. The "idols of the den," to use Bacon's phrase, must be abolished. Time will permit only a few illustrations of the kind of important variables largely overlooked or neglected by learning theorists. These variables are largely cultural, but they condition learning as that term is used in education in important ways.

One of the first of this kind of variables to come to mind is social class membership. The recent findings of the Chicago group of the differential cultural loading of group mental tests[10] favoring in general the middle class is a case in point.

At Purdue University for more than a decade through the *Purdue Opinion Panel* we have been making studies particularly of the attitudes of teen-agers in secondary schools and have massive evidence that socio-economic status is an important variable related to a large variety of learned attitudes. One of our most recent studies[11,12] based on a representative national sample shows beyond peradventure the differential incidence of deferred gratification among social classes. Middle class persons on the average conform more closely to the deferred gratification pattern than do lower class persons. It may readily be hypothesized that this difference in turn is related to a differential incidence of mental disease.

Social class differences in attitudes toward the educational

---

[9] John Dewey. *Is a Science of Education Possible?*

[10] Kenneth W. Eells and Others. *Intelligence and Cultural Differences: A Study of Cultural Learning and Problem-Solving.* Chicago: University of Chicago Press, 1951.

[11] Sverre Lysgaard. *Social Stratification and the Deferred Gratification Pattern.* Unpublished doctoral dissertation, Purdue University, August, 1952.

[12] H. H. Remmers, R. E. Horton, and Sverre Lysgaard. "Teen-Age Personalty in our Culture," Poll Report No. 32, *Purdue Opinion Panel,* Vol. 11, No. 3, May, 1952.

curriculum are already clearly apparent at the age of nine or ten. While about twenty per cent of all children dislike the various school subjects, there is a definite correlation between socio-economic status and these attitudes.[13,14] It is obvious that 'what' is learned and 'how' it is learned is importantly conditioned by social class status. And this statement has, in terms of our results from the *Purdue Opinion Panel*, broad generality. From highly personal matters such as dating to the broadest problems of international relations the evidence shows correlation of attitudes in these areas with socio-economic status.

Bennett[15] has made a significant contribution to social-psychological theory by showing that in teen-agers perceptions of the rewards and punishments imposed by the environment are a function of felt need for competence. In a competitive culture such as ours this finding has highly important consequences for both practice and theory.

Religious beliefs of teen-agers have been shown by Myers[16] to be significantly related to academic interests. Orthodox students who believe religious faith is better than logic for solving problems tend to choose courses that will not challenge their religious dogmas, are more superstitious than students with more secular orientation, have less knowledge of general facts and current events, and are significantly more frequently in the lower socio-economic status group.

Our Jewish sample is without exception better informed and typically optimistic and on the 'liberal' side. There is a possibility that we have a biased sample of pupils who are proud of their Jewish culture and therefore identify themselves as Jewish, while other Jewish children may tend to conceal their Jewish affiliation.

The persistent political literary myth that the least favored

---

[13] H. H. Remmers and R. H. Bauernfeind. *Manual, The SRA Junior Inventory.* Chicago: Science Research Associates, 1951.

[14] H. H. Remmers and B. Shimberg. *Manual, The SRA Youth Inventory.* Chicago: Science Research Associates, 1949.

[15] Edward A. Bennett. "Socio-psychological Interaction," *Further Studies in Attitudes,* Series XIX, Purdue University, September, 1951, pp. 1–36.

[16] M. Scott Myers. "The Latent Rôle of Religious Orientation," *Further Studies in Attitudes,* Series XIX, Purdue University, September, 1951, pp. 61–94.

group economically is also the most radical, not to say revolutionary, gets no support from our results. It is typically the upper group that is most in favor of so-called 'liberal' changes.

In an interesting study of the attitudes of the summer faculty of a large university and using as criterion measures the F-Scale[17] and Thurstone's Scale to Measure Attitude toward the Church,[18] Streuning[19] obtained a correlation of .42 between scores on these two measures, i.e., an authoritarian personality pattern predicts a favorable attitude toward the church and vice versa.

Differential academic achievement of the sexes has been well established by educational psychologists. It seems a plausible hypothesis, in the absence of any known differences in mental ability between the sexes, that these differentials are among other things a function of the differences in self-perceived social rôles. Such a hypothesis is particularly compelling with respect to measured knowledge of and attitude toward the specifics of child management. At least by the age of twelve to fourteen there are significant differences between boys and girls in this regard, and the gap increases with increase in age. In general girls are the better child psychologists.[20] A factorial analysis of the data also showed clearly that the criterion variable is positively correlated with age, maturation, and educational influence.

Another important variable is parental education. This is significantly related to a large variety of the attitudes and this again conditions 'what' is learned and 'how' it is learned. Incidentally, another hardy literary myth concerning the eternal opposition of youth and crabbed old age gets no support from our findings. In general on a large variety of issues the attitudes of parents can be predicted from their children's attitudes more accurately than the predictions of scholastic achievement from so-called intelligence tests.

---

[17] Adorno and Others. *The Authoritarian Personality.* New York: Harper & Brothers, 1951, pp. 255–257.

[18] L. L. Thurstone and E. J. Chave. *Measurement of Attitude toward the Church.* Chicago: University of Chicago Press, 1929.

[19] E. A. Streuning. *A Study of Certain Social and Religious Attitudes of University Faculty Members.* Unpublished M.S. thesis, Purdue University, 1951.

[20] H. H. Remmers and A. J. Drucker. "Teen-Agers' attitudes toward problems of child management." *Jour. educ. Psychol.*, February, 1951, pp. 105–113.

'Race' obviously is another important variable that has its impact on learning, conditioning as it does the individual's learning of his social rôle. This and the variables I have already mentioned make many of the variables reported in the psychological literature appear relatively trivial—serial order, backward association, reminiscence, oblivescence, stimulus error, stimulus generalization, and the like. Possibly future efforts will enable fitting the kinds of variables that I have been discussing into a theory of learning at a higher level of abstraction and generality arrived at inductively. This seems more promising to me than the hope that the 'systems' with which we now plague our graduate students will generate experimental data in their support or refutation.

To summarize: I have tried to show that learning theory as we now have it is a semantic misnomer—an animal with chameleon-like characteristics depending for its appearance upon where you find it. I have tried to suggest that what is likely to prove more fruitful is a crossbreeding of psychology with other disciplines. The well-known biological fact of the greater hardiness of hybrids than pure strains is, I believe, an appropriate metaphor in the present context. And, finally, the animal that will result from these efforts may turn out to be an elephant not white, but a powerful creature to serve the purposes of education and applied psychology generally.

A COMPARISON BETWEEN CHILDREN WHO HAVE MOVED FROM SCHOOL TO SCHOOL WITH THOSE WHO HAVE BEEN IN CONTINUOUS RESIDENCE ON VARIOUS FACTORS OF ADJUSTMENT

N. M. DOWNIE*
Purdue University

This is a study of boys and girls in one of the 'boom-towns' in the Pacific North-West. In the late 1930's Hermiston, Oregon, was a small village located in the Umatilla Desert in northeastern Oregon. In 1940 an experimental irrigation project was established there, and in the years 1941–42, 600 million dollars were allocated for the establishment of an ammunition storage depot. McNary Dam on the Columbia River is being constructed a few miles to the north.

In 1940, 306 children attended the elementary schools; in 1950 there were 1141. In 1950 there were 4500 people in the school district and estimates show that there may be 8000 by 1955. At the present time the school population is made up approximately equally of children (1) whose parents either farm or are in business in Hermiston; (2) whose parents work at the ordnance depot; (3) and of those whose fathers are construction workers at McNary Dam.

In 1949 in some classes in the Hermiston schools nineteen out of twenty-five students had attended schools other than the local schools. It has been frequently suggested that moving about from one community to another is apt to have an adverse effect on a child's intelligence test scores and his social adjustment. The research in this area is quite limited. Smith (*3*) showed that there was a slight but insignificant relationship between intelligence test scores and the number of places in which a person had lived. This study was conducted on college students. Beach and Beach (*1*) reported a study made on transient families in California in which the IQ's were shown to be the same as average. In the same study the children's adjustment to other

* The writer is indebted to Mr. Clifford C. Norris of the Hermiston, Oregon, schools for permission to use the data in this study.

pupils, toward the school, and toward the teacher was rated by the teachers. The majority of these transient children were rated as 'normal' by their teachers, about twenty per cent as possessing 'fair' adjustment, and about fifteen per cent as being 'poorly adjusted.' These children were not compared with children who had not moved. Many earlier studies have shown that children who move frequently tend to be retarded academically. This is to be expected when one considers the different procedures followed in different schools, the absence of records at the time of transfer, and the time lost in moving about.

In the fall of 1949 all children in eighteen classes making up the fifth, sixth, seventh, and eighth grades were given the intermediate form of the *Otis Self-Administering Test of Mental Ability.* During the months between the beginning of the school term and the Christmas holidays, a large amount of sociometric data was collected on each student. Each student was asked to name three other children in the class with whom he would prefer to carry out various classroom activities. Each choice was related to a real activity—action followed the choosing. Altogether twelve sociometric tests were given each student. The results of these tests, on the basis of the number of times chosen, were converted into social acceptance scores. In this study the lower the social acceptance score, the better the acceptance of the individual by the peer group.

RESULTS

In Table 1 are shown the mean scores on the *Otis Self-Administering Test of Mental Ability, Intermediate Form.* Students in the first category have attended only Hermiston schools. Because of the small numbers in the higher intervals it was felt desirable to lump all cases who had attended from six to ten schools into one group.

As a result of applying student's t-test as a test of significance, it was found that none of the differences among the various groups was significant. It is also interesting to notice that the means cluster around the mean of 100 which Otis reported for his standardization groups (*2*).

An analysis of the data by grades also showed no significant differences nor trends. In the eighth grades the means for the movers were higher than those of the non-movers. In the

TABLE 1.—NUMBER OF SCHOOLS ATTENDED AND SCORES ON THE OTIS SELF-ADMINISTERING TEST OF MENTAL ABILITY

No. Schools Attended	N	M	SD	SE_M
1	166	101.8	13.04	1.02
2	78	104.4	11.15	1.27
3	79	103.1	12.15	1.38
4	55	101.6	11.85	1.61
5	39	101.1	12.75	2.07
6–10	34	101.7	12.20	1.77

seventh grades this pattern was reversed. In the sixth grades those children who had moved one or two times had a higher average score than those who had been in Hermiston all of their academic lives or who had three or more moves. In the fifth grades the group with three or more moves had the lowest mean score.

The results from the sociometric tests are shown in Table 2.

TABLE 2.—MONTHS IN THE HERMISTON PUBLIC SCHOOLS AND SOCIAL DISTANCE SCORES

Number of Months	N	M	SD	SE_M
Over 36	43	1.85	.52	.080
25–36	41	1.54*	.51	.081
13–24	43	1.57*	.52	.080
4–12	75	1.69	.55	.063
1–3	84	1.79	.72	.078
Non-Transfers	164	1.84	.70	.055

* Significantly different from the top and bottom categories at the one per cent level of confidence.

When one remembers that the lower the score, the greater the degree of social acceptance, Table 2 shows some interesting results. At the bottom of the table is the average score of one hundred sixty-four boys and girls who have always been in the same school system. Their average score is about the same as that of transfers who have been in the Hermiston system for thirty-six or more months. The average score for transfer students who had been in the schools for periods of one or two years was significantly lower (one per cent level of confidence) than that

of those who had been in attendance longer. The average scores of students who had been in attendance less than a year were lower than those of three or more years' attendance, but not significantly so.

Table 3 shows the results of analyzing the social distance scores on the basis of the number of times the children have moved. From this table it seems that children who have moved once or twice (been in two or three schools) tend to be chosen more in their groups than children who have been in the same system continuously or who have moved three or more times. An analysis of these social distance scores by grades showed no significant differences among the groups with the four grades.

TABLE 3.—NUMBER OF SCHOOLS ATTENDED AND SOCIAL ACCEPTANCE SCORES

No. of Schools	N	M	SD
1 (Hermiston)	164	1.84	.70
2–3	160	1.65*	.56
4–10	130	1.73	.63

* Significantly different from the first category at the 5% level.

SUMMARY

1) Children in Hermiston who moved about a good deal made intelligence test scores comparable to children who had been in continuous residence in Hermiston.

2) As far as social acceptance as measured by a simple sociometric technique is concerned, the picture is more confused. This study showed that one or two moves or being in a school system from one to three years after moving seems to lead to greater average social acceptance than having been in the school throughout one's entire academic life, having moved around quite a bit, or having been in the school system less than a year.

REFERENCES

1) A. Beach and W. Beach, "Family migratoriness and child behavior." *Sociology and Social Research*, 20: 503–23, 1937.

2) S. A. Otis, *Self Administering Test of Mental Ability, Intermediate Form.* Yonkers, N. Y.: World Book Co., 1928.

3) M. Smith, "Some relationship between intelligence and geographic mobility," *American Sociological Review*, 8: 657–666, 1943.

A SIMPLIFIED ITEM ANALYSIS CARD FOR HIGH-SCHOOL AND COLLEGE INSTRUCTORS

WILLIAM C. BUDD

Goucher College

One of the reasons few instructors make little if any systematic attempt to improve their objective test items, even if they are familiar with item analysis techniques, is that they have no convenient way of recording and filing the items and the item analysis data. The writing or typing of items on ordinary filing cards is a laborious procedure if one does not have the services of a secretary. In addition, the use of the ordinary filing card for recording the item makes no provision for the inclusion of any except the minimum amount of information with respect to the analysis of the item and this usually in no special form. If the information for the items in any single test is recorded on one large sheet of paper, it is difficult to pool the items and the information from two or more separate tests.

To alleviate these difficulties, the item analysis card shown in Figure 1 has been developed and used by the author in his work at Goucher College. The preferred method of making the card is to use a stencil and a blank 5″ × 8″ filing card although simply using a stencil and one-half of an ordinary 8½″ × 11″ sheet of paper works acceptably. After an item has been administered to a group of students, the analysis is made and the data are recorded on the card. The item is then simply cut out of a copy of the test and stapled or glued to the item analysis card. This obviates the necessity for typing the item on the card. The card works especially well with multiple-choice items although it can be used for true-false items by using response 1 for true and response 2 for false. Space is provided for two separate analyses if the item is to be administered again. If certain changes are to be made in the item after any administration, these can be conveniently written on the back of the card.

The average high-school or college instructor has neither the time nor the need for a complex and rigorous method of item analysis. This particular card was designed for an analysis

using the upper and lower twenty-seven per cent of the class. On the card 'N' and '%' represent respectively the number and percentage in the high and low group giving each response. The term '% correct' refers to the actual percentage of correct responses in both groups when corrected for guessing by some formula such as $P_c = P_r - \frac{P_w}{N-1}$ where P_c is the true percentage of correct responses, P_r is the percentage of right responses,

FIGURE 1

Attach Item Here

Date: Class:					Date: Class:				
Response	High		Low		Response	High		Low	
	N	%	N	%		N	%	N	%
1.	9	69	7	55	1.				
2.	1	8	2	15	2.				
3.	2	15	2	15	3.				
4.	1	8	2	15	4.				
5.	0	0	0	0	5.				
Omit					Omit				
% Correct		61		44	% Correct				
Difficulty		53			Difficulty				
Discrimination		r = .15			Discrimination				

Suggestions on back.

P_w is the percentage of wrong responses, and n is the number of choices in the item. The index of difficulty can be obtained simply by averaging the '% correct' for the high and low group. For the index of discrimination, one of the most simple and yet thoroughly satisfactory methods is to employ the approximations to the product moment coefficient between the item and the criterion given by Flanagan.[1] To enter Flanagan's table, all

[1] John C. Flanagan, "General considerations in the selection of test items and a short method of estimating the product-moment coefficient from data at the tails of the distribution," *Jour. educ. Psychol.*, Vol. xxx, December, 1939, pp. 674–680.

that is needed is the percentage of correct responses in both the high and low groups.

Figure 1 gives an example of the use of this card for a group of fifty students. The upper and lower thirteen papers have been selected for the analysis. The percentage of students choosing each response and the corrected percentage of right responses for both groups have been calculated. Using the latter figures, the indices of difficulty and discrimination have been found. Some notation about the advisability of changing response 5 would probably be made on the back of the card.

The exact form of the card is not to be considered fixed. Individual instructors may wish to change portions of it to fit their own peculiar needs. The utility of the card does not lie in the particular method of item analysis employed but simply in the convenient method it provides for recording and filing the data on individual test items.

BOOK REVIEWS

HAROLD G. SHANE AND E. T. MCSWAIN. *Evaluation and the Elementary Curriculum.* New York: Henry Holt and Company, 1951.

Shane and McSwain have given us a book we had to have, *Evaluation and the Elementary Curriculum.* If curriculum development is to take place primarily at the classroom level—and it must—then the means of judging the actual quality of the learning experiences that take place must also be available at that level. Our thinking about curriculum-making has moved a long way since the twenties, when the course of study approach to curriculum-making was at its zenith; this book is an essential step in the modern curriculum movement.

The authors accept the local school unit as the basis for curriculum-making, and this is fundamental. If they did not, the approach to evaluation they use—indeed, evaluation itself—would make little sense.

What they have attempted is this: to present a non-technical discussion of the nature of evaluation, and the associated problems of value formation and value conflict; then to apply the rationale suggested by this discussion to the elementary curriculum as it is usually found in the schools. Their hope is that " . . . all persons concerned with the educational program share in designing it . . . " by means of systematic, widespread evaluation.

Now let us see how the authors went about their task. The text of the book is divided into two parts: Part I, "Educational Values and the Elementary Curriculum," and Part II, "What Evaluation Suggests for Curriculum Practices." Part I occupies 123 pages; Part II occupies 288 pages. The bulk of the book is devoted to attempts to apply the principles that are developed in Part I. Following Part II are three appendixes: "A Summary of the Development of Evaluation in Education" (a documented account of the growth of evaluation from its parents, philosophy and science); "Criteria of Good Citizenship in Behavioral Terms" (an example of the crucial step in evaluation—translating goals from abstract to behavioral terms); and "An Annotated Bibliography of Evaluation Instruments and Related

Materials." This last appendix is the consequence of an extensive first-hand survey of the instruments actually in use in major current evaluation projects, as well as a survey of the considerable literature of evaluation. The appendix includes, also, an extensive annotated bibliography of books and articles.

This is a book in which the appendix material is fully as important as the textual material; the third appendix, especially (on materials, instruments, and literature) is itself a contribution to the literature.

We have a book, then, with three parts: theory, application, and tools. Let us see how it came off.

Remembering that the book "has been planned as a practical resource for teachers, administrators, and parents" (p. vii), one would first of all ask that it be so written as to be useful and understandable to those people, who, above, everything, plead with the education writer: "lead us not into confusion."

Everything considered, the book passes this test. The topics undertaken are plainly set forth, from (for example) "Why Elementary Education Needs Evaluation," through chapters on the 3 R's and Social Studies, to the leadership problems associated with coöperative evaluation. Section by section, the topics move logically from one to another, and some of them are given novel and refreshing analyses. Chapter 2, "Values To Be Sought with Children in the Elementary School," is one of these.

The authors faced a significant problem in their attempt to write on this topic to their announced audience; how comprehensive should they be, and how detailed, to make the application of evaluation clear, yet avoid spreading their discussion over too much ground? Unless the ideas are applied, they have little meaning. Should the authors attempt to cover many topics briefly, or go more deeply into a more restricted number of topics?

The authors chose to be comprehensive, yet to keep the book of moderate size. There is almost nothing about the elementary school that isn't at least touched on in this book. Often, the thumbnail sketch of an educational problem is fresh and illuminating as is the case in Chapter 15, "Applying Educational Values to Some Persistent Elementary School Problems." Sometimes the sketch is too brief, as happened in the page-and-a-half discussion of workshops; or the discussion would benefit from

better chosen detail, as is the case in the discussion of grouping and promotion.

If there is a difference of any importance between the judgment of the authors and this reviewer, it is here—the book seems too comprehensive. The authors appear to have been too strongly attracted by the desirability of including material usually found in textbooks on the elementary curriculum. They review the history of a trend, (e.g., the various grouping plans, pp. 293–302) at greater length than is required to advance the argument; they include lists of experiences children might have in the school subjects, in one or two cases (this reviewer thinks) somewhat uncritically.

However, this criticism will function as a recommendation, for some readers. Shane and McSwain have succeeded in giving us an excellent source book. If what one wants is (among other things) a running start on the myriad of details, this book offers it by virtue of its very comprehensiveness.

For this reason, as well as others—its novelty and its generally careful documentation and extensive use of research—the book will probably find its way into many a college class on elementary curriculum. This is as it should be. The authors' realistic approach to the school should help the pre-service teacher. The careful scholarship should help the more advanced student. The comprehensiveness will make the book of value as a manual to the teachers, administrators, and parents facing actual school problems.

Evaluation and the Elementary Curriculum is, on the whole, as good a statement of its kind as could now be made. The difficulty the authors faced, and dealt with as well as the present state of our understanding permits, is the same one that has plagued educators from the beginning: how can we be sure that human, moral values penetrate the entire school curriculum? Shane and McSwain suggest that we can now go much farther in this direction than many a formalist thinks. The day will come, it is to be hoped, when we can go the whole way. ARTHUR W. FOSHAY
The Ohio State University

DOROTHY YOST DEEGAN. *The Stereotype of the Single Woman in American Novels.* New York: King's Crown Press, Columbia University, 1951, pp. 252.

The fact that large numbers of women in America do not or cannot marry presents a problem with both sociological and psychological implications. Personal adjustment, social attitudes, education and other factors are involved. In the study of these problems, whatever the approach, the element of the social attitude, as expressed in the stereotype of the single woman, will be an important factor. Novel-length fiction probably offers the best literary media for detailed study of this stereotype. This treatise reports such a study with a discussion of the implications for society and for education.

In the investigation, single women are defined as women thirty years of age or older whose probability of marriage is slight. On the basis of content, the study was limited to the one hundred twenty-five novels of the Dickinson best-novel list "in which the setting, action and characters are all wholly within the boundaries of the United States and are contemporary with the author's lifetime." Careful reading of the novels provided information for a case study of each woman character. Classification categories provided material on childhood and adolescence, ambition and achievement, human relationships, attitudes, and factors in nonmarriage and adjustment.

Analysis of the classification of personnel factors revealed a recurring pattern. The most impressive positive element in the pattern is the tendency to stereotype. Elements most clearly evident refer to vocation. More than half had no gainful vocation. The majority of the working women were school teachers, dressmakers or domestics. Among others, two distinct patterns that emerged were gossipmongers and maiden sisters. On the negative side were found discrepancies between the fiction portrait of single women and the facts found in investigations as in psychosexual studies, and studies of eminent women. Comparison of a larger range of American fiction with the sampling revealed some of the same pattern with considerable change in details. For instance, there is less discrepancy between fiction and fact. The persistent social attitude toward the single woman is reflected by the repeated implication that her work is not satisfying nor wholly commendable. The attitude is far more derogatory than otherwise.

Implications for society and education are briefly discussed. If single women are not so efficient nor so well adjusted as they

should be, one reason for this is the social attitude toward them. Society has much to learn about this problem and it "can be taught." The author advocates "that society should educate women to face the possibility of singleness. . . . "

This study was well conceived and ably conducted. The findings are of considerable importance to the sociologist and the psychologist. However, since an important objective is to make applications of the implications for society and for education, it is unfortunate that there was not greater emphasis upon attitudes toward the single woman found in more recent novels. Both the status of the single woman and attitudes toward her have changed appreciably during the last generation.

MILES A. TINKER

University of Minnesota

LOUIS KAPLAN AND DENIS BARON. *Mental Hygiene and Life.* New York: Harper and Brothers, 1950, pp. 442.

In fifteen chapters the following subjects are discussed: mental illness in American life, the mental hygiene movement, the emergence of personality, personality in human relationships, 'psychic forces' in human behavior, how needs influence behavior, man and his emotions, emotional development and personality, emotional development and mental health, how frustrations and conflicts influence our lives, utilizing stress situations in the growth process, how we adjust to tension, the deterioration of personality, and you and your mental health.

The book is planned with two central themes, one "that there can be no wholesome, rational, moral society until we can develop wholesome, rational, moral individuals," and, two, "That the development of such individuals is a process which every person must make his own responsibility."

The style is planned to be as simple as possible, with a minimum of technical terms, and with illustrations to help the 'ordinary' reader. The simple style which "The professional may find . . . to be an oversimplified version of human behavior," is justified in terms of the need for everyone, at least "every educated person" to know the principles of mental hygiene.

There is a brief summary concerning the great amount of mental illness, suggestions concerning the curability of mental

illness, and an explanation of the meaning and scope of The Mental Hygiene Movement.

The basic psychology of the text is a modified Freudian psychoanalysis. "Fifty years have so modified Freud's pronouncements that we now refer to this science as 'psychodynamics' to distinguish it from the earlier theories which drew forth such bitter criticisms." (p. 118) There are the conscious mind, the unconscious mind, and the subconscious mind. Emphasis is placed upon the statement that the mind is not a thing, it "is a symbol of mental activity." (p. 122) And the psychic forces, "the Id, the Ego, and the Superego . . . are not structures or entities. They are simply symbolic terms which are used to describe certain emotional and psychological activities . . . " (pp. 122–23). For a comprehensive statement as to the problems involved, the reader is referred to references cited.

The main topics are emotion and personality, with emphasis upon adjustment, conflict, child problems, frustration, needs, stress and tension, the self, especially the self-concept, and more on prevention than the index suggests.

The general method of treatment emphasizes the emergence of the personality in its development, particularly, of the self-concept and in connection with social stimuli, and increasingly more complex responses to more and more complex situations. Personality is defined as "the individual's interpretation of himself in relation to the environment as this is expressed in behavior." (p. 69) The discussions are more helpful than the definition.

There are discussions of many problems of daily life; and helpful explanations are offered for certain of the problems of human behavior, either with the help of or in spite of the theoretical psychological assumptions. More psychological information is brought into the text than is indicated by the index, e.g., discussions of habit and of prevention.

Many psychologists will not be willing to accept the idea that the psychoanalytic psychology of the present is as free from adverse criticism as the authors seem to think, or that it is the only kind of 'psychodynamics.' None, on the other hand, who is informed, will deny the great and significant contributions that have been made by several schools of psychoanalysis.

The student who desires a thorough and balanced understanding of the problems discussed, will need to do extensive reading

and study of some of the other and excellent references given in this text. The volume will probably find extended use especially as a text for short courses.

A. S. EDWARDS

The University of Georgia.

MARGARET NAUMBURG. *Schizophrenic Art: Its Meaning in Psychotherapy.* New York: Grune & Stratton, Inc., 1950, pp. 247.

The spontaneous art productions of two schizophrenic girls, one aged eighteen and the other twenty-five, represent the source material for this attractively put-together volume. Art expression of these two subjects is presented in detail; the emergence and unfolding of their psychodynamics are interestingly and vividly illustrated. The clinical material is presented to throw light on diagnosis as well as therapy; however its main objective seems to be the use of art expression for therapy. The book contains a survey of the literature on neurotic and psychotic art from 1876 to 1950; an appendix of the poems and notes of the two girls whose art productions are illustrated and analyzed; and a brief index. There are sixty-three illustrations and four color plates. All in all, the book is in itself an art production and is a more clinically illustrated follow-up of the artist's first volume on art expression.

The present volume as well as the original, called *Studies of the 'Free' Art Expression of Behavior Problem Children and Adolescents as a Means of Diagnosis and Therapy,* are both the outgrowth of research projects in art therapy developed in the New York State Psychiatric Institute. A review of the first volume appeared in this JOURNAL (Vol. 39, No. 6, October, 1948, pp. 382–384). What is said in the first review is equally applicable to this present volume. The author's verbalizations not only reveal methods and attitudes which can be helpful to the therapist in encouraging release and making more resourceful use of material as well as developing possibilities for obtaining more valuable material which can be of help to both diagnosis and therapy. This is illustrated step by step in detail, so the unfolding process is much more vividly portrayed here.

The volume will be of greatest use, of course, to the professionals in the field, particularly those who are interested in the use

of art expression in therapy. But the general discussion of psychodynamics of schizophrenia as illustrated in art and therapy will be of interest to all therapists whether or not they use art therapy. The material reads well enough so that it can be of interest to people interested in art and human nature though not professionals. For the professional field this represents a valuable contribution to humanized therapy.

H. MELTZER

Psychological Service Center
St. Louis, Missouri